# Chapter 1

# Preliminaries

## 1 Sets

Sets are the building blocks of modern mathematics. Almost all mathematical objects can be described in terms of sets and relations between them. In this section we will introduce the basic concepts from set theory.

**(1.1)** Naively speaking a *set* is just an *unordered collection of objects*. We can describe a set by putting its objects between braces. E.g. the set of all vowels is

$$V := \{a, e, i, o, u\}.$$

A set is an unordered collection, so we can also write

$$V := \{e, u, a, i, o\}.$$

An object can occur only once in a set so $\{a, a\}$ is the same as $\{a\}$. A set can also contain an infinite number of objects. An example of this is the set of natural numbers

$$\mathbb{N} := \{0, 1, 2, 3, 4, 5, \dots\}.$$

**(1.2)** To express that a certain object is in a given set or not we can use the mathematical symbols $\in$ and $\notin$

$$a \in V \text{ but } b \notin V$$

We say that $a$ is an element of $V$.
Two sets are the same if they contain the same elements:

$$V = W \iff \forall x : (x \in V \iff x \in W)$$

**(1.3)** Another way of describing a set is by putting between braces the defining property of its elements

$$V := \{x | x \text{ is a vowel}\} \text{ and } \mathbb{N} := \{x | x \text{ is a natural number}\}$$

However defining a set in this way can be tricky: sometimes you end up defining something that cannot be a set. An example of this is the collection of all sets that do not contain themselves. If this collection were a set we run into problems: does this set contain itself or not? To exclude these contradictory sets one needs to work out an axiomatic theory for sets. This was done by Zermelo and Fraenckel and standard set theory is still refered to ZermeloFraenkel set theory with the axiom of choice.

**(1.4)** A special set is the *empty set*. It contains no objects and it is denoted by $\{\}$ or $\emptyset$. By the previous definition of equality of sets, it is easy to see that there is just one unique empty set.
A set $B$ is called a *subset* of $A$ if all elements in $B$ are also in $A$:

$$B \subset A \iff \forall x \in B : x \in A.$$

The empty set is a subset of all sets including itself. In general a set with $n$ elements has $2^n$ subsets. F.i. the subsets of $\{1, 2\}$ are

$$\emptyset, \{1\}, \{2\}, \{1, 2\}$$

The set containing all the subsets of a given set $A$ is called the *power set* of $A$

$$\mathcal{P}(A) := \{x | x \subset A\}$$

So

$$\mathcal{P}(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$$

**(1.5)** Given two sets $A$ and $B$ one can define the *union* $A \cup B$ as the set that contains elements of $A$ and $B$.

$$A \cup B = \{x | x \in A \text{ or } x \in B\}$$

The *intersection* $A \cap B$ is the set of elements that are both in $A$ and $B$

$$A \cap B = \{x | x \in A \text{ and } x \in B\}$$

Two sets are called *disjoint* if there intersection is the empty set.
The *difference* $A \setminus B$ is the set of elements that are in $A$ but not in $B$

$$A \setminus B = \{x | x \in A \text{ and } x \notin B\}$$

**(1.6) Aside.**
The power of the concept of a set is that a set can be an element of another set. F.i. the set

$$\{\{\}, \{\{\}\}\}$$

is a set with two elements: the empty set and the set that contains the emptyset. In principle one could build the whole of mathematics starting from the empty set and constructing new sets like the one above. An example of this are the natural numbers: one defines

$$\begin{aligned} 0 &:= \emptyset \\ 1 &:= \{0\} = \{\emptyset\} \\ 2 &:= \{0, 1\} = \{\emptyset, \{\emptyset\}\} \\ &\vdots \\ N + 1 &:= N \cup \{N\}. \end{aligned}$$

In this way the set $N$ contains exactly $N$ elements. Following this framework we get that

$$\mathbb{N} := \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \dots\}$$

More elaborate constructions allow us to construct from this the set of integers $\mathbb{Z}$, the set of rational numbers $\mathbb{Q}$, the set of real numbers $\mathbb{R}$ and the set of complex numbers $\mathbb{C}$, but we will not pursue this direction. We can consider these sets as subsets of each other in the following way:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

# 2 Relations

**(1.7)** A *couple* is an ordered pair of two objects (those objects can be the same). We denote the couple consisting of two objects $a$ and $b$ by $(a, b)$. By definition $(a, b) \neq (b, a)$. In a similar way we can define triples, quadruples etc.
The *Cartesian product* of two sets $A$ and $B$ is the set consisting of all couples with the first element in $A$ and the second element in $B$.

$$A \times B := \{(a, b) | a \in A, b \in B\}$$

E.g.

$$\begin{aligned} &\{0, 1\} \times \{2, 3, 4\} = \{(0, 2), (0, 3), (0, 4), (1, 2), (1, 3), (1, 4)\} \text{ but} \\ &\{2, 3, 4\} \times \{0, 1\} = \{(2, 0), (3, 0), (4, 0), (2, 1), (3, 1), (4, 1)\} \end{aligned}$$

The cartesian product of a set with $n$ elements with a set with $m$ elements has $nm$ elements and therefore we speak of a product.

**(1.8)** A *relation* $\bowtie$ between two sets $A$ and $B$ is a subset of their cartesian product: $\bowtie \subset A \times B$. We call $A$ the *source* and $B$ the *target* of $R$.
Given a relation $\bowtie$ we write $a \bowtie b$ if $(a, b) \in \bowtie$ and $a \not\bowtie b$ if not. The *domain* of $\bowtie$ is the subset of the source containing all elements that occur in a couple of $R$. The *image* is the subset of the target containing all elements that occur in a couple of $\bowtie$.

$$\mathrm{Dom} \bowtie := \{a \in A | \exists x \in B : a \bowtie x\} \text{ and } \mathrm{Im} \bowtie := \{b \in B | \exists x \in A : x \bowtie b\}$$

In general relations become only interesting if they satisfy some extra properties. In the next subsections we will have a look at some special types of relations.

## 2.1 Maps

**(1.9)** A relation $f \subset A \times B$ is called a map if for every element $a$ in the source there is a unique $b$ in the target such that $afb$:

$$\forall a \in A : \exists! b \in B : afb.$$

In that case we denote this unique $b$ by $f(a)$. The map itself is sometimes represented by $f : A \to B : a \longmapsto f(a)$.

**(1.10)** A map is called

- *surjective* if every $b$ in the target is in the image of $f$.

  $$\forall b \in B : \exists a \in A : b = f(a)$$

  The map

  $$\pi : \mathbb{Z} \to \mathbb{N} : a \longmapsto |a|$$

  is an example of a surjective map.

- *injective* if different elements in $A$ are mapped to different elements in $B$

  $$\forall x, y \in A : x \neq y \implies f(x) \neq f(y)$$

  The map

  $$\iota : \mathbb{Z} \to \mathbb{Z} : a \longmapsto 2a$$

  is an example of an injective map.

- *bijective* if it is injective and surjective.

  $$\forall b \in B : \exists! a \in A : b = f(a)$$

  A bijective map from a set to itself is called a *permutation*.

The map

$$\beta : \mathbb{Z} \to \mathbb{N} : a \longmapsto \begin{cases} 2a & a \geq 0 \\ -1 - 2a & a < 0 \end{cases}$$

is an example of a bijective map.

**(1.11)** Two maps $f : A \to B$ and $g : B \to C$ can be *composed* to obtain a new map

$$g \circ f : A \to C : a \longmapsto g(f(a)).$$

For any set $A$ we can define the *identity map* as

$$\mathbb{I}_A : A \to A : a \longmapsto a$$

These maps have the special property that for any other map $f : A \to B$ we get

$$\mathbb{I}_B \circ f = f = f \circ \mathbb{I}_A$$

**(1.12)** For any relation $R$ we can define the *inverse relation* $R^{-1}$ as the relation that switches the role of source and target:

$$R^{-1} = \{(b, a) | (a, b) \in R\} \subset B \times A.$$

**(1.13) Lemma.**
*The inverse of a map is again a map if and only if the map is bijective.*

*Proof.* the inverse of a map $f$ is also a map if for every element $b$ in the target there is a unique element in the source such that $b = f(a)$. This implies that $f$ is both injective and surjective. □

**(1.14) Aside. The axiom of choice**
Apart from the obvious axioms for set theory, the standard theory includes an axiom that is a little bit controversial: the axiom of choice. This axiom states that for any set $S$ the elements of which are all nonempty sets, there exists a map

$$f : S \to \bigcup_{X \in S} X$$

such that $f(X) \in X$. Such a map is called a choice function, because it allows you to choose one element from each $X$. The axiom used to be controversial in the beginning of the 20th century because it merely states the existence of such a function, but it does not indicate how one can construct

such a function. For finite sets one can prove the axiom from the other axioms of set theory but for infinite sets this is generally not possible.
Nowadays the axiom of choice is accepted by most mathematicians, because it is needed in many important results in mathematics. In linear algebra it is needed to prove the existence of a bases for all vector spaces.

## 2.2 Operations and groups

**(1.15)** An *operation* on a set $A$ is a map

$$\diamondsuit : A \times A \to A.$$

For an operator we write $a\diamondsuit b$ for the image of $(a, b)$ under $\diamondsuit$.
Operations become more interesting if they have special properties:

A Associativity:

$$\forall a, b, c \in A : a\diamondsuit(b\diamondsuit c) = (a\diamondsuit b)\diamondsuit c$$

This property enables us to leave the brackets out and write $a\diamondsuit b\diamondsuit c$ without confusion.

N The existence of an neutral element:

$$\exists e \in A : \forall a \in A : a\diamondsuit e = e\diamondsuit a = a$$

I The existence of inverses:

$$\forall a \in A : \exists \mathrm{inv}(a) \in A : a\diamondsuit \mathrm{inv}(a) = \mathrm{inv}(a)\diamondsuit a = e$$

C Commutativity:

$$\forall a, b \in A : a\diamondsuit b = b\diamondsuit a$$

**(1.16)** A set $A$ equipped with an operation $\diamondsuit$ that satisfies the first three properties (A,N,I) is called a *group*. If it satisfies also the commutativity, we call the group commutative or abelian.
Groups occur a lot in mathematics. The standard example is the addition for integers

$$+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} : (a, b) \longmapsto a + b$$

with as neutral element $0$ and as inverse $\mathrm{inv}(a) = -a$. It is clear that this group is commutative.

**(1.17) Example.**
Another basic example is the group of permutations on a set

$$\mathrm{Perm}(A) := \{\phi : A \to A | \phi \text{ is a permutation}\}$$

equipped with the composition as operation:

$$\circ : \mathrm{Perm}(A) \times \mathrm{Perm}(A) \to \mathrm{Perm}(A) : (\phi, \psi) \to \phi \circ \psi$$

In general when $A$ has more than two elements, this group is not commutative.

**(1.18) Aside.**
Many groups occur as groups of permutations with special properties.

- The group of isometries of the Euclidean plane $\mathbb{E}$ is the set of permutations that preserve the distance between points.

$$\mathrm{Isom}(\mathbb{E}) := \{\phi \in \mathrm{Perm}(\mathbb{E}) | \forall x, y \in \mathbb{E} : d(\phi(x), \phi(y)) = d(x, y)\}$$

- The group of translations of the Euclidean plane $\mathbb{E}$ is the set of isometries that map line to parallel lines and fix either no or all points.

$$\mathrm{Trans}(\mathbb{E}) := \{\phi \in \mathrm{Isom}(\mathbb{E}) | \forall x, y \in \mathbb{E} : \phi(x)\phi(y) // xy \& \phi(x) = x \implies \phi(y) = y\}$$

## 2.3 Equivalence relations

Now we take a look at relations with the same source and target.

**(1.19)** A relation $R \subset A \times A$ is called

R *reflexive* if

$$\forall a \in A : aRa.$$

S *symmetric* if $R = R^{-1}$ or

$$\forall a, b \in A : aRb \iff bRa.$$

T *transitive* if

$$\forall a, b, c \in A : aRb \& bRc \implies aRc$$

If a relation $\simeq$ satisfies these three properties then we will call it an equivalence relation.

**(1.20)** Equivalence relations are very important tools in mathematics. Many objects in mathematics are not exactly the same but very alike. The standard way to expres this, is by introducing an equivalence relation. The three main properties of an equivalence relation formalize this alikeness: an object is alike itself (reflexivity), if A is like B then by is like A (symmetry) and if A is like B and B is like C then A is like C (transitivity).

**(1.21)** For any element $a \in A$ we can define the subset consisting of all elements equivalent to $a$:

$$[\![a]\!] := \{b | b \simeq a\}.$$

This set is called the *equivalence class* of $a$.
We can also construct the set of all equivalence classes in $A$

$$A/\simeq := \{[\![a]\!] | a \in A\}$$

This set is called the *quotient set*. There is also a natural *quotient map*

$$\pi : A \to A/\simeq : a \longmapsto [\![a]\!]$$

This is clearly a surjective map.

**(1.22)** A *partition* of a set $A$ is a set $B \subset \mathcal{P}(A)$ such that $\forall U, V \in B : U \neq V \iff U \cap V = \emptyset$ and $\cup_{x \in B}$. In words, it's a set of disjoint nonempty subsets of $A$ such that their union is $A$. It is clear from its construction that for any equivalence relation $\simeq$ the quotient set $A/\simeq$ is a partion of $A$.
For any partition $B$ one can define an equivalence relation on $A$: $x \simeq_B y \iff \exists U \in B : x \in U \& y \in U$. Reflexivity and symmetry follow directly from the definition, transitivity follows from the fact that the sets in $B$ are disjoint. The quotient of this relation $A/\simeq_B$ is $B$. Vice versa if one starts from a partition that is the quotient of an equivalence relation $\simeq$ then the equivalence relation $\simeq_{A/\simeq}$ is the same as the original. We can conclude that partitions and equivalence relations are two ways of talking about the same concept.

**(1.23) Example.**
If we start from a surjective map $f : A \to B$ we can also construct an equivalence relation $a \simeq_f b \iff f(a) = f(b)$. There is a bijection between $A/\simeq_f$ and $B$ that maps $[\![a]\!]$ to $f(a)$.

**(1.24) Example.**
Consider the set $\mathcal{L}$ of lines in the Euclidean plane. The relation we consider

is parallelism: $u//v$ if and only if $u$ and $v$ are parallel lines. It is easy to check that this is an equivalence relation. The quotient set $\mathcal{L}///$ contains all parallel classes of lines. There is a nice bijection between $\mathcal{L}///$ and $\mathbb{R} \cup \{\infty\}$ that associates to each parallel class the slope of its lines.

**(1.25) Example.**
Consider the set of integers and choose an integer $p$. Define $\simeq_p := \{(a, b) | a - b$ is a mulitple of $p\}$. This is an equivalence relation because $0$ is a multiple of $p$ (R), $-x$ is a multiple of $p$ iff $x$ is (S) and the sum of two multiples of $p$ is again a multiple of $p$ (T).
The quotient set $\mathbb{Z}_p := \mathbb{Z}/\simeq_p$ consists of $p$ elements: for every $0 \leq i < p$ the set of all integers that have rest $i$ after division by $p$ is an equivalence class.

**(1.26) Example.**
Consider the set $\mathcal{T}$ of all triangles in the plane. Two triangles are equivalent if there is an isometry (translation, rotation, reflexion) of the plane that transforms the one into the other. This is an equivalence relation because the identity is an isometry (R), the inverse of an isometry is an isometry (S) and the composition of two isometries is an isometry (T).
The quotient set $\mathcal{T}/\simeq$ consists of the congruence classes of triangles Note that there is a nice bijective map from $\mathcal{T}/\simeq$ to

$$\{(a, b, c) \in \mathbb{R}^3 | a \geq b \geq c \& b + c > a\}$$

**(1.27) Example.**
Consider the set $\mathbb{E} \times \mathbb{E}$ of couples of points in the Euclidean plane. We say that $(x, y) \cong (z, w)$ if there is a translation $t : \mathbb{E} \to \mathbb{E}$ such that

$$(t(x), t(y)) = (z, w)$$

This is an equivalence relation because the identity is a translation (R), the inverse of a translation is a translation (S) and the composition of two translations is a translation (T).

**(1.28) Example.**
More general consider a set $A$ and a set $G \subset \mathrm{Perm}(A)$ that forms a group under composition. We say that $x \simeq y$ if there exists a $g \in G$ such that $g(x) = y$. Analogously to the previous examples one can show that $\simeq$ is an equivalence relation. The equivalence classes are in this case are also called orbits and the quotient set is denoted by $A/G$.

**(1.29) Example.**
If $S$ is a set with an equivalence relation $\cong$ and $T \subset S$ then $\cong_T := \cong \cap T \times T$ is also an equivalence relation. The quotient $T/\cong_T$ can be mapped to $S/\cong$ by

$$\iota : T/\cong_T \to S/\cong: \{y \in T | y \cong_T x\} \longmapsto \{y \in S | y \cong x\}$$

and this map is an injection.

## 2.4 Partial orders

A final type of relations that frequently occurs are partial orders

**(1.30)** A relation $\prec \subset S \times S$ is called a *partial order* if it is reflexive, transitive and antisymmetric:

$$\forall x, y \in S : x \prec y \& y \prec x \implies x = y.$$

$S$ is called a partially ordered set or poset.

**(1.31) Examples.**

The set $\mathbb{N}$ equiped with the relation smaller than or equal $\leq$ is a partial order (as is $\geq$)

The set $\mathbb{N}$ equiped with the relation $a|b$ ($a$ divides $b$).

The set $\mathcal{P}(S)$ equiped with the relation $\subset$.

If $\prec$ is a partial order then its inverse relation $\prec^{-1}$ is also a partial order.

If $S$ is a set with a partial order $\prec$ and $T \subset S$ then $\prec_T := \prec \cap T \times T$ is also a partial order.

**(1.32)** Given a partial order $\prec \subset S \times S$ An element $s \in S$ is called *maximal* if there is no $y \in S$ such that $x \prec y$ and $x \neq y$.
An element $s \in S$ is called *minimal* if there is no $y \in S$ such that $y \prec x$ and $x \neq y$.

**(1.33) Examples.**

The set $\mathbb{N}$ equiped with the relation smaller than or equal has a minimal element $0$.

The set $\{x \in \mathbb{N} | x \geq 2\}$ equiped with the relation $a|b$ has as minimal elements all prime numbers and it has no maximal element.

The set $\mathcal{P}(S)$ equiped with the relation $\subset$ has a minimal element, the empty set, and a maximal element, $S$.

If $\prec$ is a partial order then its maximal elements are the minimal elements of its inverse relation and vice versa.

**(1.34)** Note that partial orders are called partial because not every two elements in the set can be compared: i.e. there can be pairs of elements $a, b$ for which neither $a \prec b$ nor $b \prec a$ (think of the relation $|$: $2 \not| 3$ and $3 \not| 2$. A partial order is called a total order if two elements are always comparable:

$$\prec \text{ is total } \iff \forall x, y \in S : x \prec y \text{ or } y \prec x.$$

It is clear that $\leq$ is a total order on $\mathbb{N}$ but $|$ is not. $\subset$ is also not total on $\mathcal{P}(S)$.

# Chapter 2

# The main characters

Linear algebra is the study of two mathematical objects, vector spaces and linear maps. In this chapter we introduce both and give lots of examples. However before starting to talk about vector spaces and linear maps we need to review some things about fields, because these play a very important role in the background. A field is an object that allows you to work with just like with the ordinary real numbers: you can do addition, substraction, multiplication and division.

## 1 Fields

**(2.1) Definition.**
A *field* is a set $\mathbb{K}$ with two operations $+ : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ and $\cdot : \mathbb{K} \times \mathbb{K} \to \mathbb{K}$ such that

- $\mathbb{K}, +$ is a commutative group. We denote the neutral element by $0$ and the 'inverse' by $-x$.
- $\mathbb{K} \setminus \{0\}$ is a commutative group. We denote the neutral element by $1$ and the inverse by $\frac{1}{x}$.
- the distributive law holds:

$$\forall x, y, z \in \mathbb{K} : x \cdot (y + z) = (x \cdot y) + (x \cdot z)$$

**(2.2) Examples.**
The standard examples are the number fields

- $\mathbb{Q}, +, \cdot,$
- $\mathbb{R}, +, \cdot,$
- $\mathbb{C}, +, \cdot.$

**(2.3)** The definition of a field can be relaxed to allow that the multiplication is not commutative (f.i. the quaternions $\mathbb{H}$) or does not have inverse elements (f.i. the integers $\mathbb{Z}$) or both (f.i. the $2 \times 2$-matrices over $\mathbb{R}$). In these cases we speak of *rings* instead of fields.

**(2.4) Example.**
A slightly more exotic example is the finite field with $p$ elements. To construct this we start from $\mathbb{Z}$ equiped with the equivalence relation $\simeq_p$ as laid out in the previous chapter. The quotient set $\mathbb{Z}_p$ consists of $p$ elements:

$$\begin{aligned}
[\![0]\!] &= \{\dots, -2p, -p, 0, p, 2p, \dots\} \\
[\![1]\!] &= \{\dots, -2p+1, -p+1, 1, p+1, 2p+1, \dots\} \\
&\vdots \\
[\![p-1]\!] &= \{\dots, -p-1, -1, p-1, 2p-1, 3p-1, \dots\} \\
[\![p]\!] &= [\![0]\!] \\
[\![p+1]\!] &= [\![1]\!] \\
&\vdots
\end{aligned}$$

We define the sum and product of classes as

$$\begin{aligned}
[\![a]\!] + [\![b]\!] &:= [\![a+b]\!] \\
[\![a]\!] \cdot [\![b]\!] &:= [\![a \cdot b]\!]
\end{aligned}$$

These definitions do not depend on the choice of $a$ and $b$ as representatives for their classes. Indeed if we add a multiple of $p$ to $a$ or $b$ the new sum and product will differ from the old by a multiple of $p$.

**(2.5) Theorem.**
*If $p$ is a prime number then $\mathbb{Z}_p, +, \cdot$ is a field.*

*Proof.* The associativity, commutativity, distributivity, the existence of neutral elements for the multiplication and the addition and the existences of inverses for the addition follow directly from the fact that these also hold in $\mathbb{Z}$. The only property that does not hold in $\mathbb{Z}$ is the existence of inverses for the multiplication.
If $p$ is a prime and $a$ is not a multiple of $p$, we know that the greatest common divisor of $a$ and $p$ is 1. The algorithm of Euclid provides us with a way to espress the gcd as a linear combination of $a$ and $p$

$$1 = ax + py \text{ with } x, y \in \mathbb{Z}$$

If we turn this into an equation of classes

$$[\![1]\!] = [\![ax + py]\!] = [\![ax]\!] = [\![a]\!] \cdot [\![x]\!]$$

we see that $[\![x]\!]$ is the multiplicative inverse of $[\![a]\!]$. □

If it is clear that we are working over $\mathbb{Z}_p$ we usually write $a$ as a shorthand for $[\![a]\!]$.

**(2.6)** For any field $\mathbb{K}$, one can consider polynomials in one variable. These are expressions of the form

$$f(X) = a_n X^n + \ldots a_1 X + a_0 \text{ with } a_i \in F.$$

The biggest $i$ for which $a_i \neq 0$ is called the degree of $f$. If the degree is $1$ then we call $f$ linear. The set of all polynomials in one variable over $\mathbb{K}$ is denoted by $\mathbb{K}[X]$.

**(2.7)** A root of $f$ is an element $u$ of the field such that $f(u) = 0$.
In a field a linear polynomial has always a unique root:

$$u = \frac{-a_0}{a_1}$$

Equations of higher degree have not neccesary solutions and the number of solutions depend on the field. For instance the equation $X^2 + 1 = 0$ has no solutions in $\mathbb{Q}$ or in $\mathbb{R}$, 2 in $\mathbb{C}$ ($i$ and $-i$) and $\mathbb{Z}_5$ ($[\![2]\!]$ and $[\![3]\!]$) and only one in $\mathbb{Z}_2$ ($[\![1]\!]$). On the other hand the equation $X^2 - X - 1 = 0$ has no solutions in $\mathbb{Q}$ or $\mathbb{Z}_2$ but 2 solutions in $\mathbb{R}$, $\mathbb{C}$ and one in $\mathbb{Z}_5$.
For a general field one can say that if $f(X)$ has $a$ as a root then $f(X) = (X - a)g(X)$ for some polynomial $g(X)$ and a polynomial of degree $n$ has at most $n$ different roots.

**(2.8)** Among the fields we have seen the field of complex numbers has a very special property: every polynomial of nonzero degree has at least one root.

$$\forall f \in \mathbb{K}[X] : \deg f > 0 \implies \exists a in \mathbb{K} : f(a) = 0.$$

Such a field is called *algebraically closed*. Algebraicly closed fields are very important because they make certain aspects of linear algebra easier.

# 2 Vector spaces

**(2.9)** Traditionally vectors were used to represent velocity, acceleration, forces and all kinds of other physical concepts that could be described as arrows in threedimensional space: quantities with both size and direction. These objects allowed certain mathematical operations: they could be added to each other and they could be scaled: multiplied by a real number.
As mathematics developed further, mathematicians began to recognize that a lot of sets of objects are closed under taking sums and scaling. Examples of these are solutions system of homogeneous linear equations, or homogeneous partial differential equations or certain sets of functions. Such sets will be called vector spaces and their elements vectors.

**(2.10)** A vector space is always defined over a field. In this course the field will almost always be $\mathbb{R}$ or $\mathbb{C}$ but sometimes we will also consider a finite field of the form $\mathbb{Z}_p$. Where we don't want to be specific about the field, we shall call it $\mathbb{K}$.

**(2.11) Definition. Vector space**
A *vector space* over a field $\mathbb{K}$ is a nonempty set $V$, equipped with operations of addition ($+$) and scalar multiplication by elements of $\mathbb{K}$ (that is such that for any two vectors $v, w \in V$ and for any scalar $\lambda \in \mathbb{K}$, the sum $v + w$ and scalar multiple $\lambda v$ are defined and in $V$),
and such that the following axioms hold.

**AG** The addition $+ : V \times V \to V$ is a commutative group structure on $V$ with neutral element $0_V$ and inverse $\mathrm{inv}(v) = -v$.

**M1 (associativity)** for all $\lambda$, $\mu$ in $k$, for all $v \in V$, $\lambda(\mu v) = (\lambda\mu)v$

**M2** for all $v \in V$, $1v = v$ (where $1 = 1_{\mathbb{K}}$ is the 1 of the field $\mathbb{K}$)

**M3 (right distributivity)** for all $v \in$V, for all $\lambda, \mu \in \mathbb{K}$, $(\lambda + \mu)v = \lambda v + \mu v$

**M4 (left distributivity)** for all $v, w \in$V, for all $\lambda \in \mathbb{K}$, $\lambda(v + w) = \lambda v + \lambda w$

The axioms simply demand that the addition and scalar multiplication in $V$ follow the standard rules of arithmetic which we are used to seeing in sets of vectors over the real numbers.
We can deduce various rules from the axioms such as

$$0_{\mathbb{K}} v = 0_v, \quad (-1)v = -v \quad \text{and} \lambda 0_v = 0_v$$

for all $v \in V$ and for all $\lambda \in \mathbb{K}$, which must then hold in any vector space (and which we shall frequently use without comment).

**(2.12) Examples.** 1. The simplest example is $V = \{0\}$ with $0 + 0 = 0$ and $\forall \lambda \in \mathbb{K} : \lambda 0 = 0$. This is called the *trivial vector space* or the *null space*.

2. We have certainly already met the vector spaces $\mathbb{R}^3$, of row vectors with 3 real coordinates, i.e.

$$\{(x, y, z) : x, y, z \in \mathbb{R}\}$$

Given vectors $(x, y, z), (x', y', z')$ we define

$$(x, y, z) + (x', y', z') = (x + x', y + y', z + z'), \quad \lambda(x, y, z) = (\lambda x, \lambda y, \lambda z)$$

The zero of this vector space is $(0, 0, 0)$ the inverse of $(x, y, z)$ is $(-x, -y, -z)$.

3. More generally, when $n$ is any positive integer and $\mathbb{K}$ any field we define the vector space $\mathbb{K}^n$ to be the set of all row vectors with $n$ coordinates each from $\mathbb{K}$, i.e.

$$\{x_1, x_2, \ldots x_n) : x_i \in \mathbb{K}\}$$

(we often call that the set of $n$-tuples over $\mathbb{K}$), where addition and scalar multiplication are defined by

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

and

$$\lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$$

In this example

$$0_V = (0_k, 0_k, \ldots, 0_k)$$

and

$$-(x_1, x_2, \ldots, x_n) = (-x_1, -x_2, \ldots, -x_n)$$

. We could also see the elements of $\mathbb{K}^n$ as column vectors; and sometimes we shall do this.

4. Similarly we can define $\mathbb{K}^{\mathbb{N}}$ to be the set of all sequences over $k$, with addition and scalar multiplication defined coordinatewise. Then $0_V$ is the sequence consisting only of zeros, and and

$$-(x_1, x_2, \ldots, ) = (-x_1, -x_2, \ldots, )$$

.

5. For any field $\mathbb{K}$, and integers $m, n$, the set of $m \times n$ matrices $(x_{ij})$ with $x_{ij}$ in $\mathbb{K}$ is a vector space when we define

$$(x_{ij}) + (y_{ij}) = (x_{ij} + y_{ij})$$

and

$$\lambda(x_{ij}) = (\lambda x_{ij})$$

(i.e. usual addition and scalar multiplication of matrices).

$0_V$ is the $m \times n$ matrix of zeros, and $-v$ is formed by negating all the entries in the matrix $v$.

6. The set of all maps $f : \mathbb{R} \to \mathbb{R}$ is a vector space, where for maps $f, g$ and $\lambda \in \mathbb{R}$ we define $f + g$ and $\lambda f$ by the rules

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x)$$

The zero map and the negative $-f$ of $f$ are defined by the rules

$$0(x) = 0 \quad (-f)(x) = -(f(x)$$

And we have another example if we replace $\mathbb{R}$ by $\mathbb{C}$ here.

7. Using the same addition and scalar multiplication we see that the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$, $C^0(\mathbb{R})$, and the set of all differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, $C^1(\mathbb{R})$, are vector spaces over $\mathbb{R}$. To verify this we need to check that when $f, g$ are continuous (or differentiable) then so is $f + g$, and so is $\lambda f$, for any $\lambda \in \mathbb{R}$.

8. The set $\mathbb{R}[x]$ of all polynomials in $x$ with real coefficients is a vector space over $\mathbb{R}$, where addition and scalar multiplication are defined by the rules

$$\begin{aligned}
&(a_0 + a_x + \ldots + a_n x^n) + (b_0 + b_1 x + \ldots + b_n x^n) \\
&= (a_0 + b_0) + (a_1 + b_1)x + \ldots + (a_n + b_n)x^n \\
&\lambda(a_0 + a_1 x + \ldots + a_n x^n) \\
&= \lambda a_0 + \lambda a_1 x + \ldots \lambda a_n x^n
\end{aligned}$$

Of course this addition and multiplication are the addition and multiplication defined above for functions.

9. The set of vectors $(x, y, z)$ that are solutions to a linear equation such as $x + y + z = 0$ is a vector space under the standard addition and multiplication of vectors in $\mathbb{R}^3$. We just have to verify that the sum of two solutions and a scalar multiple of a solution are also solutions.

More generally the set of vectors $(x_1, \ldots x_n)$ that are solutions to a system of linear equations

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n &= 0 \\
a_{31}x_1 + a_{32}x_2 + \ldots a_{3n}x_n &= 0 \\
\vdots \quad \vdots \quad \vdots &= 0 \\
a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n &= 0
\end{aligned}$$

is a vector space

10. The set of functions $f$ that are solutions to the differential equation

$$f"(x) + 3f'(x) + 2f(x) = 0$$

is a vector space over $\mathbb{R}$ (also over $\mathbb{C}$). We need to verify that the sum of two solutions is a solution, and that a scalar multiple of a solution is a solution. This is easy to verify and is a consequence of the linearity of the equation.

11. The set $\mathcal{P}(\Omega)$ of all subsets of a set $\Omega$, where the sum $A+B$ of two sets $A, B$ is defined to be its symmetric difference, that is

$$A + B = (A \cup B) \setminus (A \cap B)$$

is a vector space over the field of two elements $\mathbb{Z}_2$, The axioms force scalar multiplication to be defined as follows:-

$$1A = A, \quad 0A = \emptyset$$

12. Let $A$ be any set and let $\mathbb{K}^A$ be the set of all maps from $A$ to $\mathbb{K}$ and define for $f, g : A \to \mathbb{K}$

$$f + g : A \to \mathbb{K} : a \longmapsto f(a) + g(a)$$
$$\lambda \cdot f : A \to \mathbb{K} : a \longmapsto \lambda f(a)$$

The set of all maps for which only a finite number of elements in $A$ map not to zero also form a vector space with the same operations.

13. Let $V$ be the set $\mathbb{E} \times \mathbb{E}/\simeq$ from example 1.27. We can turn this into a vector space over $\mathbb{R}$ if we define the addition by

$$[\![(p, q)]\!] + [\![(q, r)]\!] = [\![(p, r)]\!]$$

(note that for every class we can choose a representative starting in a given point $q$) and scalar multiplication by

$$\lambda[\![(p, q)]\!] = [\![(\phi_\lambda p, \phi_\lambda q)]\!]$$

where $\phi_\lambda$ is a scaling with factor $\lambda$.

14. Let $V = \{x \in \mathbb{R} | x > 0\}$ be the set of positive real numbers. We turn this into a vector space over $\mathbb{R}$ by putting

$$x + y := xy \text{ and } \lambda \cdot x := x^\lambda.$$

15. If $V$ and $W$ are two vector spaces, we define

$$V \oplus W := V \times W \text{ with}$$
$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2) \text{ and}$$
$$\lambda(v, w) = (\lambda v, \lambda w)$$

$V \oplus W$ is a vector space and it is called the direct sum of $V$ and $W$.

**(2.13) Aside.**
We end this section with some near misses.

- The set $V = \mathbb{R}^2$ with the standard vector addition and scalar multiplication defined as,

$$\lambda(x_1, x_2) = (x_1, \lambda x_2)$$

is NOT a vector space. One can check that all axioms hold except M3:

$$(x_1, (\lambda + \mu)x_2) \neq (x_1, \lambda x_2) + (x_1, \mu x_2) = (2x_1, (\lambda + \mu)x_2).$$

If we define

$$\lambda(x_1, x_2) = (0, \lambda x_2)$$

then all axioms hold except M2.

- Another way that $V$ can fail to be a vector space is although all axioms hold, the object $\mathbb{K}$ is not a field. If $\mathbb{K}$ is still a ring then we call $V$ a module. Modules are very close to vector spaces but there are some big structural differences too.

The additive group $\mathbb{Z}_p, +$ can be considered as a module over $\mathbb{Z}$ by putting $n \cdot x = x + \cdots + x$ ($n$ times). This is an example of a module that unlike a vector space does not have a basis (see chapter **??**).

The space $\mathrm{Mat}_{2\times 1}(\mathbb{R})$ of real $2 \times 1$-matrices can be seen as a module over the ring of real $2 \times 2$-matrices where we use the standard matrix addition and multiplication. (Note that $\mathrm{Mat}_{2\times 1}$ is also a vector space over $\mathbb{R}$)

# 3 Linear maps

**(2.14)** Let $V$ and $W$ two vector spaces over the field $\mathbb{K}$. A map $\phi : V \to W$ is called a *linear map* if it is compatible with the two operations:

(i) $\forall x, y \in V : \phi(x + y) = \phi(x) + \phi(y)$

(ii) $\forall x \in V : \forall \lambda \in \mathbb{K} : \phi(\lambda x) = \lambda \phi(x)$

We can pack condition (i) and (ii) together in one condition:

(*) $\forall x, y \in V : \forall \lambda, \mu \in \mathbb{K} : \phi(\lambda x + \mu y) = \lambda\phi(x) + \mu\phi(y)$

**(2.15) Examples.** 1. Let $V$ be the vector space of all $3 \times 1$-matrices, $W$ be the vector space of $4 \times 1$ matrices and $A$ be any $4 \times 3$-matrix. The map

$$\phi_A : V \to W : x \to Ax$$

is a linear map.

2. The map $\frac{d}{dx} : C^1(\mathbb{R}) \to C^0(\mathbb{R}) : f \longmapsto \frac{df}{dx}$ is linear.
3. Let $g$ be any continuous function. The map $R_g : C^0(\mathbb{R}) \to C^0(\mathbb{R}) : f \longmapsto f \circ g$ is linear. on the other hand the $R_g : C^0(\mathbb{R}) \to C^0(\mathbb{R}) : f \longmapsto g \circ f$ is only linear if $g(x) = \lambda x$ for some $\lambda \in \mathbb{R}$
4. The map zero map $0 : V \to W : x \longmapsto 0$ is linear and a constant map is linear if and only if it is the zero map.
5. Let $S$ be a set and $s \in S$. The map

$$\iota_s : \mathcal{P}(S) \to \mathbb{Z}_2 : X \longmapsto \begin{cases} 1 & s \in X \\ 0 & s \notin X \end{cases}$$

and the map

$$\nu : \mathcal{P}(S) \to \mathbb{Z}_2 : X \longmapsto \begin{cases} 1 & \#X \text{is odd} \\ 0 & \#X \text{is even} \end{cases}$$

are linear.

6. For any subset $Y \subset S$ the map $\mathcal{P}(S) \to \mathcal{P}(S) : X \longmapsto X \cap Y$ is linear but the map $\mathcal{P}(S) \to \mathcal{P}(S) : X \longmapsto X \cup Y$ is not.
7. $V$ be the set $\mathbb{E} \times \mathbb{E}/\simeq$ from **??**. If $\phi$ is an isometry of the plane then

$$V \to V : [\![(x, y)]\!] \longmapsto [\![(\phi(x), \phi(y))]\!]$$

is linear.

8.
9. If $V$ and $W$ are two vector spaces and $f : V_1 \to V_2$ and $g : W_1 \to W_2$ are two linear maps. We define

$$f \oplus g : V_1 \oplus W_1 \to V_2 \oplus W_2 : (v, w) \longmapsto (f(v) \oplus f(w)).$$

This linear map is called the direct sum of $f$ and $g$.

**(2.16)** The set of all linear maps from $V$ to $W$ is denoted by $\mathrm{Hom}(V, W)$. $\mathrm{Hom}(V, W)$ is itself a vector space. We can add two linear maps $\phi$ and $\psi$

$$\phi + \psi : V \to W : x \longmapsto \phi(x) + \psi(x)$$

and multiply a map with a scalar.

$$\lambda\phi : V \to W : x \longmapsto \lambda\phi(x)$$

Both maps are linear (take care for the second map one needs de commutativity of the multiplication in the field, so this does not hold for modules) and one can easily check the axioms for a vector space.

**(2.17)** If we have two linear maps $\phi : V \to W$ and $\psi : W \to Z$ then the composition

$$\psi \circ \phi : V \to Z : x \longmapsto \psi(\phi(x))$$

is also a linear map.
If a linear map is a bijection then the inverse of this map is also linear:

$$\phi(\lambda\phi^{-1}(x) + \mu\phi^{-1}(y)) = \lambda\phi(\phi^{-1}(x)) + \mu\phi(\phi^{-1}(y)) = \lambda x + \mu y$$

so

$$\lambda\phi^{-1}(x) + \mu\phi^{-1}(y) = \psi^{-1}(\lambda x + \mu y).$$

For any vector space $V$ the set of linear bijections is denoted by $GL(V)$. The composition of two elements in $GL(V)$ is again an element of $GL(V)$ and also the inverse of an element is in $GL(V)$. Therefore $GL(V)$ is a group (the neutral element is the identity element and the associativity follows because the composition of maps is always associative). This group describes all symmetries of the vector space $V$. Unlike $\mathrm{Hom}(V, V)$ the set $GL(V)$ is not a vector space.

## 4 The main problems of linear algebra

We now have defined two major objects vector spaces and linear maps. A natural question one can ask is then can we describe or classify these objects.
Describing all vector spaces is a hopeless business as there are so many. Luckily this is not neccesary because many vector spaces look alike: they have the same structure, only their elements have different names.
To make the concept of vector spaces that look alike more formal we introduce the concept of isomorphic vector spaces

**(2.18) Definition.**
We will call two vector spaces *isomorphic* if there is a bijective linear map between them.

$$V \cong W \iff \exists \phi \in \mathrm{Hom}(V, W) : \phi \text{ is a bijection.}$$

Any linear bijection is also called an isomorphism.

**(2.19)** If we have a set of vector spaces [1] then the relation $\cong$ gives an equivalence relation on this set.

- Reflexivity: the identity map is a linear bijection.
- Symmetry: the inverse of a linear bijection is again a linear bijection.
- Transitivity: the composition of two linear bijection is again a linear bijection.

**(2.20) Examples.**

The vector space $\mathbb{R}^3$ and the vector space $V = \{a_0 + a_1X + a_2X^2 | a_i \in \mathbb{R}^3\}$ are isomorphic because the map

$$\phi : \mathbb{R}^3 \longmapsto V : (\lambda, \mu, \nu) \longmapsto \lambda + \mu X + \nu X^2$$

is a linear bijection.
The vector space $C_0^1 := \{f \in \mathbb{C}^1(X) | f(0) = 0\}$ is isomorphic with $C^0(X)$ because the linear map

$$\frac{d}{dt} : C_0^1 \to C^0 : f \to \frac{d}{f} dt$$

has an inverse

$$\int : C^0 \to C_0^1 : f \to \int_0^x f(t)dt.$$

The vector space $\mathrm{Maps}(S, \mathbb{Z}_2)$ and the vector space $\mathcal{P}(S)$ are isomorphic because

$$\phi : \mathrm{Maps}(S, \mathbb{Z}_2) \to \mathcal{P}(S) : f \longmapsto \{x \in S | f(x) = 1\}$$

is linear with inverse

$$\phi^{-1} : \mathcal{P}(S) \to \mathrm{Maps}(S, \mathbb{Z}_2) : X \longmapsto f \text{ such that } f(x) = 1 \iff x \in X.$$

[1] We have to specify a set of vector spaces because the collection of all vector spaces is not a set

The real solutions $V$ of the linear equation $x + y + z = 0$ and the real solutions $W$ of the differential equation $f\text{“} + f = 0$ are isomorphic by the linear map

$$\phi : W \to V : f \longmapsto (f(0), f'(0), f''(0) - f'(0))$$

the inverse of which is

$$\phi^{-1} : V \to W : (a, b, c) \longmapsto a\cos t + b\sin t$$

The main questions now become:

**(2.21) Question.**
*Can we classify the isomorphism classes of vector spaces and can we find for every class a nice representative?*

**(2.22) Question.**
*Given two nice representatives of isomorphism classes of vector spaces can we describe the linear maps between them?*

# Chapter 3

# Finite Dimensional Vector Spaces

## 1 Subspaces and spans

Some of the examples we have given of vector spaces live inside other vector spaces.

**(3.1) Definition.**
Let $V$ be a vector space over a field $\mathbb{K}$. A subset $X \subset V$ is called a subspace if $X$ is a vector space with the addition and scalar multiplication are inherited from $V$.

If one wants to find out whether a certain nonempty set is indeed a subspace, one first needs to check whether the addition and scalar multiplication are well defined on $X$:

(i) $\forall x, y \in X : x + y \in X$,

(ii) $\forall \lambda \in \mathbb{K} : \forall x \in X : \lambda x \in X$,

Once these two conditions are checked the axioms M1-M4 are automatically fullfilled in $X$ because they are fullfilled in $V$. To check the group axioms [AG] we have to be a bit more carefull: associativity and commutativity follow directly from $V$ and we can infer that the neutral element and inverse elements are in $X$ because we can write them as $0 \cdot x$ and $-1 \cdot x$ and then use $(ii)$.
Conditions (i) and (ii) together are equivalent to the condition

$$\forall v, w \in X : \forall \lambda, \mu \in \mathbb{K} : \lambda v + \mu w \in X. \quad (*)$$

so a nonempty subset $X \subset V$ is a subspace if and only if $(*)$ holds.

**(3.2) Examples.** 1. For any vector space $V$, $\{0\}$ and $V$ itself are both subspaces.

2. The set of solutions $(x, y, z)$ to the equation $x+y+z = 0$ is a subspace of $\mathbb{R}^3$, and in general so is the set of solutions to a set of homogenous linear equations (see examples 2.12).

3. The spaces $C^0(\mathbb{R})$ of all continuous functions, all differentiable functions, all polynomial functions $\mathbb{R}([x])$ with coefficients in $\mathbb{R}$ are subspaces of the space of all functions from $\mathbb{R}$ to $\mathbb{R}$. The space of all solutions to $f''(x) + 3f'(x) + 2f(x) = 0$ is a subspace of the space of all functions from $\mathbb{R}$ to $\mathbb{R}$ (or from $\mathbb{C}$ to $\mathbb{C}$)..

4. For any integer $n$ the set of polynomials of degree at most $n$ is a subspace of $\mathbb{R}[x]$, the space of all polynomials.

5. The set of all functions of the form $a\sin(x + b)$ with $a, b \in \mathbb{R}$ form a subspace of $C^0(\mathbb{R})$.

6. For any vectors $v$, $w$ in a vector space $V$ over $k$,

$$\{\lambda v : \lambda \in k\} \quad \text{and} \quad \{\lambda v + \mu w : \lambda, \mu \in k\}$$

are subspaces of $V$. These two examples are part of a larger example which we shall reach soon.

7. Consider the vector space $\mathcal{P}(S)$ over $\mathbb{Z}_2$. The set of subsets of $S$ with an even number of elements is a subspace. The subset of all sets where the number of elements is a multiple of 3 is not a subspace.

Given two subspaces of a vector space, their intersection and sum (not union) always gives us two new ones.

**(3.3) Proposition.**
*Suppose that $V$ is a vector space over $k$ and $X$ and $Y$ are subspaces of $V$. Then*

**(i)** *$X \cap Y$ is also a subspace of $V$ (and of both $X$ and $Y$).*

**(ii)** *where $X + Y = \{x + y : x \in X, y \in Y\}$, $X + Y$ is a subspace of $V$*

*Proof.* Both the intersection aand sum contain $0$ so we only need to check (*). For the intersection this follows from the fact that (*) holds for $X$ and $Y$ so $\lambda u + \mu v$ sits in $X$ and $Y$ and hence in the intersection.
For the sum we see that if $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ then

$$\lambda(x_1 + y_1) + \mu(x_2 + y_2) = \underbrace{(\lambda x_1 + \mu x_2)}_{\in X} + \underbrace{(\lambda y_1 + \mu y_2)}_{\in Y} \in X + Y$$

□

If we look at the direct sum we see that the vector spaces $(V, 0)$ and$(0, W)$ are subspaces of $V \oplus W$. It is clear that their intersection is the null space $(0, 0)$.

We can generalise the last of our examples from 3.2

**(3.4) Definition. Linear combinations, spans**
Let $V$ be a vector space (over $\mathbb{K}$). Any vector $\lambda_1 v_1 + \lambda_2 v_2 + \ldots \lambda_m v_m$ is called a linear combination of $v_1, v_2, \ldots v_m \in V$.
For any subset $X$ the set of all linear combinations of elements of $X$

$$\langle X \rangle = \{\sum \lambda_i v_i : \lambda_i \in \mathbb{K} \& v_i \in X\}$$

is called the span of $X$. For the emptyset we define $\langle \emptyset \rangle = \{0\}$.

We can also write $\langle \{v_1, \ldots, v_m\} \rangle$ as $\langle v_1, \ldots, v_m \rangle$, that is, the brackets $\{, \}$ are not necessary.

**(3.5) Proposition.**
*$\langle X \rangle$ is a subspace of $V$. Indeed it is the smallest subspace containing $X$.*

*Proof.* To prove the first statement we check condition (*). To prove the second, we simply have to verify that any subspace containing $X$ must contain $\lambda_1 v_1 + \ldots + \lambda_t v_t$, for any $t \in \mathbb{N}$, which can be proved by induction on $t$. □

**(3.6) Examples.**
1. In $\mathbb{R}^3$, define $e_1$ to be the vector $(1, 0, 0)$, $e_2$ to be the vector $(0, 1, 0)$. Then $\langle e_1, e_2 \rangle$ is the set of all vectors of the form $(x, y, 0)$, which we often call the $xy$-plane.
2. In $k[x]$, $\langle 1, x, x^2 \rangle$ is the subspace of all polynomials of degree at most 2, and $\langle x, x^2 \rangle$ is the subspace of all polynomials of degree at most 2 that take the value $0$ at $0$.
3. In $k^n$ let $e_i$ be the vector with a $1$ as its $i$-th coordinate and a $0$ everywhere else, so that

   $$e_1 = (1, 0, 0 \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots e_n = (0, 0, 0, \ldots, 0, 1).$$

   The $\langle e_1, e_2, \ldots, e_n \rangle$ is the whole of $k^n$.
4. Let $P \subset \mathcal{P}(S)$ the set of all subsets of $S$ with 2 elements. The span of $P$ is the subspace containing all the sets with an even number of elements

5. If $v \in V$ then we will also write $\mathbb{K}v$ instead of $\langle v \rangle$ because $\langle v \rangle := \{\lambda v | \lambda \in \mathbb{K}\}$

6. Consider the vector space $\mathbb{K}^{\mathbb{N}}$ and the elements $e_i$ which are zero everywhere except on the $i^{th}$ entry where there is a one. The span $\langle e_1, e_2, \ldots \rangle$ is not equal to $\mathbb{K}^{\mathbb{N}}$ but it is the subspace of vectors with only a finite number of nonzero entries.

**(3.7) Proposition.**
*If $S$ and $T$ are subsets of $V$ then*

- $\langle S \cup T \rangle = \langle S \rangle + \langle \mathsf{T} >$
- $\langle S \cap T \rangle \subset \langle S \rangle \cap \langle \mathsf{T} >$ *(but they are not always equal)*
- $\langle \langle S \rangle \rangle = \langle S \rangle$
- $\langle S \rangle = S$ *if and only if $S$ is a subspace.*

**(3.8) Definition. Spanning sets**
Suppose that $U$ is a subspace of a vector space $V$. A set $X$ of vectors whose span is equal to $U$ is called a spanning set for $U$. We also say that $U$ is spanned by $X$.

Looking at the examples above,we see that $\{e_1, e_2\}$ is a spanning set for the $xy$-plane, $\{1, x, x^2\}$ is a spanning set for the space of polynomials of degree at most 2, $\{e_1, e_2, \ldots e_n\}$ is a spanning set for $k^n$.
Every vector space has a spanning set: the vector space itself. But this is spanning set is not interesting. The interesting spanning sets are those that are as small as possible.

**(3.9) Definition.**
A spanning set $S$ of a vector space $V$ is minimal if and only if no subset $T \subset S$ is a spanning set of $V$.

Note that the set of spanning set of $V$ forms a partially ordered set with $\subset$. The minimal spanning sets are the minimal elements of this poset.
A minimal spanning set for a vector space has some very special properties. In order to study these we need to look at the concept of linear dependence.

# 2 Linear dependence

**(3.10) Definition.**
Suppose that $V$ is a vector space over $\mathbb{K}$ We say that the set of vectors $\{v_1, v_2, \ldots, v_m\}$ in $V$ is linearly dependent if for some $\lambda_1, \lambda_2, \ldots, \lambda_m$, not all zero,

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m = 0.$$

We say that the set of vectors is linearly independent if it is not linearly dependent.

Notice that the empty set is linearly independent by definition. An infinite set of vectors is called linearly independent if every finite subset is linearly independent.

**(3.11) Proposition.**
*Suppose that $v_1, v_2, \ldots, v_m$ are vectors in a vector space $V$. Then $\{v_1, v_2, \ldots, v_m\}$ is linearly independent if and only if any vector $v$ in $\langle v_1, v_2, \ldots, v_m\rangle$ has a unique representation as a linear combination*

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m$$

*Proof.* We'll prove the equivalent result that the vectors $v_1, v_2, \ldots, v_m$ are linearly dependent if and only if there is at least one vector in $\langle v_1, \ldots, v_m\rangle$ with more than one representation as a linear combination of those vectors. Linear dependence implies that the zero vector has at least 2 representations: i.e. all $\alpha_i$ are zero or the coefficients coming from the linear combination that gives zero.
On the other hand suppose that $w = \lambda_1 v_1 + \cdots + \lambda_m v_m = \mu_1 v_1 + \cdots + \mu_m v_m$ with at least one $\lambda_i \neq \mu_i$ then

$$0 = (\lambda_1 - \mu_1)v_1 + \cdots + (\lambda_m - \mu_m)v_m$$

at least one of the $(\lambda_i - \mu_i)$ is nonzero so the $v_1, \ldots, v_m$ are linearly dependent. □

Proposition 3.11 explains why linearly independent sets of vectors are useful. We have already met some very natural examples of linearly independent sets.

**(3.12) Examples.** 1. In $\mathbb{K}^n$, $e_1, e_2, \ldots, e_m$ are linearly independent for all $m \leq n$. ( Recall that we defined $e_i$ to be the vector with its $\mathbb{K}$-th coordinate equal to 1, and all other entries 0.)

For $\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_m e_m = (\lambda_1, \ldots, \lambda_m)$

2. In $\mathbb{R}^3$, $e_1 = (1,0,0)$, $e_1 + e_2 = (1,1,0)$ and $e_1 + e_2 + e_3 = (1,1,1)$ are linearly independent. As the system of linear equations

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 0 \\ \alpha_2 + \alpha_3 &= 0 \\ \alpha_3 &= 0 \end{aligned}$$

3. $1, x, x^2, x^3, \ldots, x^n$ are linearly independent in $\mathbb{K}[x]$.

4. If $X_1, \ldots, X_n$ are disjoint subsets of $S$ then they are linearly independent in $\mathcal{P}(S)$.

5. If $S$ has an odd number of elements then the sets $S \setminus \{x\}$ where $x \in S$ are linearly independent, if $S$ has an even number of elements then

6. $\{sin(kx) | k \in \mathbb{N}\}$ are linearly independent in $C^1(\mathbb{R})$. Indeed if $f(x) = \sum_{k \leq N} \lambda_k \sin(kx) = 0$ then

$$\int_0^{2\pi} \sin(\ell x) f(x) = \frac{\lambda_\ell}{\pi} = 0$$

7. Choose $\{u_1, \ldots, u_n\} \subset \mathbb{R}$ and let $f_i(X) = \frac{(X-u_1)\ldots(X-u_n)}{(}X - u_i)$. These $f_i$ are linearly independent. Indeed if $f(X) = \lambda_1 f_1 + \lambda_n f_n = 0$ then $f(u_i) = \frac{(u_i-u_1)\ldots(u_n-u_n)}{(}u_i - u_i)\lambda_i = 0$ and as the factor before $\lambda_i$ is nonzero, $\lambda_i$ must be zero.

**(3.13) Proposition.**
*A set $S$ is linearly independent if and only if $S$ is a minimal spanning set of $\langle S \rangle$.*

*Proof.* If $S$ were not minimal, then there is a $v \in S$ such that $v \in \langle S \setminus \{v\} \rangle$ so we can write $v = \lambda_1 w_1 + \cdots + \lambda_m w_m$ with $w_i \in S \setminus \{v\}$. The vectors in $S$ are linear dependent because $\lambda_1 w_1 + \cdots + \lambda_m w_m + (-1)v = 0$
Vice versa suppose that $\lambda_1 v_1 + \cdots + \lambda_m v_m = 0$ for some $v_i \in S$ with $\lambda_j \neq 0$ then $v_j \in \langle S \setminus \{v_j\} \rangle$ because

$$v_j = \frac{-\lambda_1}{\lambda_j} v_1 + \cdots + \frac{-\lambda_m}{\lambda_j} v_m.$$

so $\langle S \setminus \{v_j\} \rangle = \langle S \rangle$ and hence $S$ is not minimal. □

**(3.14) Aside.**
By this theorem a minimal spanning set and a linear independent spanning

set are the same thing. This is only the case for vector spaces, not for modules over a ring. The reason for this is that in a ring one cannot always find an inverse for $\lambda_j$. For instance the $\mathbb{Z}$-module $\mathbb{Z}_{15}$ has minimal spanning sets $\{1\}$ and $\{3, 5\}$ but neither set is a basis because in $\mathbb{Z}_{15}$ we have that

$$0 = 15 \cdot [\![1]\!] = 5 \cdot [\![3]\!] + 3 \cdot [\![5]\!].$$

So the sets are not linearly independent. In this example there are no nonempty linearly independent subsets of $\mathbb{Z}_{15}$.

## 3 Bases

**(3.15) Definition.**
If $V$ is a vector space over $\mathbb{K}$, a basis (or base) for $V$ is a linearly independent spanning set of $V$.

NB. The plural of basis is bases - one basis, two bases.

**(3.16) Definition.**
The dimension of a vector space $V$, $\dim(V)$ is defined to be the minimal size of a basis, if $V$ has a finite basis, otherwise it is infinite.

In fact if $V$ has a finite basis, then every basis has the same size. We'll prove that soon in theorem 3.22, but aren't quite ready to do that yet.

**(3.17) Examples.**
$\{e_1, \ldots, e_n\}$ is a basis for $\mathbb{K}^n$, so is $\{e_1, e_1+e_2, e_1+e_2+e_3, \ldots\ e_1+e_2+\ldots e_n\}$. Theorem 3.22 will tell us this space is $n$-dimensional.
$1, x, x^2, \ldots$ is a basis for $\mathbb{K}[x]$ and $1, x, x^2, \ldots, x^n$ a basis for the subspace of $\mathbb{K}[x]$ consisting of all polynomials of degree at most $n$. Theorem 3.22 will tell us this space is $n+1$-dimensional.
$\{e^x, e^{2x}\}$ is a basis for the set of solutions to the differential equation

$$f''(x) - 3f'(x) + 2f(x) = 0.$$

$\{e^{ix}, e^{-ix}\}$ and $\{\cos(x), \sin(x)\}$ are two different bases for the set of solutions to the differential equation

$$f''(x) + f(x) = 0.$$

Theorem 3.22 will tell us this subspace of the space of all functions from $\mathbb{C}$ to $\mathbb{C}$ is 2-dimensional.

Both $\{(1,0,2,1),\ (0,1,-6,-1),\ (0,0,9,5)\}$ and $\{(1,0,2,1),\ (3,1,0,2),\ (0,1,3,4),\}$ are bases for the subspace of $\mathbb{R}^4$ spanned by $\{(1,0,2,1),\ (3,1,0,2),\ (0,1,3,4),\ (5,1,4,4),\ (5,0,1,$ (see examples **??**). Theorem 3.22 will tell us this subspace of $\mathbb{R}^4$ is 3-dimensional.

If $v_1,\ldots,v_n \subset V$ is a basis for $V$ and $w_1,\ldots,w_m$ is a basis for $W$ then $(v_1,0),\ldots,(v_n,0),(0,w_1),\ldots,(0,w_m)$ is a basis for $V \oplus W$.

**(3.18) Proposition.**
*Any finite spanning set for a vector space $V$ contains a basis.*

*Proof.* Let $S_0 = \{v_1, v_2, \ldots, v_n\}$ be a spanning set for $V$. If it is not minimal then there is a $v_i \in S$ such that $v_i \in \langle S \setminus \{v_i\}\rangle$. We remove $v_i$ from $S$ to obtain a new spanning set. If this new set is not minimal we again remove an element. Because $S$ is finite we can only do this a finite number of times until we get a minimal spanning set which is a basis. □

The following result is tremendously useful, with lots of consequences.
In this section we show that every linearly independent subset of a finite dimensional vector space can be extended to a basis.

**(3.19) Lemma. The exchange lemma**
*Let $A = \{x_1, x_2, \ldots x_r\}$ be a finite linearly independent set and $C = \{y_1, y_2, \ldots, y_s\}$ a finite spanning set in a vector space $V$. Then there is a subset $C'$ of $C$ such that $A \cup C'$ is a spanning set of $V$, of the same size as $C$.*

*Proof.* We construct $C'$ in $r$ steps, as the last in a sequence of sets

$$C_0 = C, C_1, \ldots C_r = C'$$

We form $C_1$ by deleting from $\{x_1\} \cup C$ the first element in $C$ (using the order given) which is in the span of its predecessors in $C$ together with $x_1$. There must be such an element by proposition **??** since $C$ spans and so $\{x_1\} \cup C$ is linearly dependent.
Then we simply iterate this process, i.e. for each $i = 1, \ldots r-1$, we form $C_{i+1}$ out of $C_i$ by deleting the first element out of $C_i$ which is in the span of its predecessors in $C_i$ together with $x_1, x_2, \ldots, x_i, x_{i+1}$. Again proposition **??** ensures we can find such an element, since $\{x_1, x_2, \ldots x_i\} \cup C_i$ is a spanning set, so $\{x_1, \ldots, x_{i+1}\} \cup C_i$ is linearly dependent; but the linear dependence of $A$ ensures that none of the $x_j$'s is in the span of $x_1, \ldots, x_{j-1}$. □

**(3.20) Corollary.**
*If $V$ is a finite dimensional vector space then any linearly independent set of vectors in $V$ can be extended to a finite basis.*

*Proof.* We just apply the exchange lemma (3.19) with $A$ as the linearly independent set, $C$ as a finite basis. □

## 3.1 Dimension-Basis Theorem

The Dimension-Basis theorem (3.22) is easily derived from a corollary of the exchange lemma.

**(3.21) Corollary.**
*Let $A$ be a finite linearly independent set and $C$ a finite spanning set of vectors in a vector space $V$. Then $|A| \leq |C|$.*

*Proof.* Let $C'$ be as in the exchange lemma. Then $A$ is a subset of $A \cup C'$, which has the same size as $C$. So $|A| \leq |A \cup C'| = |C|$. □

**(3.22) Theorem. Dimension-Basis theorem**
*If $V$ is a finite dimensional vector space, every basis contains the same number of elements.*

*Proof.* Let $B$ be a finite basis, and $B'$ a second basis. Then $B'$ must also be finite. For any finite subset $A$ of $B'$ is linearly independent, and corollary 3.21 then tells us that $|A| \leq |B|$. But an infinite set can't have a bound on the size of its finite subsets. So $B'$ is finite, and in that case, by corollary 3.21, $|B'| \leq |B|$.
Now applying the same argument to $B'$ and $B$ (rather than $B$ and $B'$) gives $|B| \leq |B'|$. □

**(3.23) Examples.**

- The dimension of $\mathbb{K}^n$ is $n$.
- The dimension of the set of solutions of $f'' - 3f' + 2f = 0$ is 2.
- The dimension of $\{Asin(x + B)\}$ is 2 because $\{sin(x), sin(x + \pi/2)\}$ is a basis.
- The dimension of $\mathcal{P}(S)$ equals the number of elements in $S$.
- The dimension of the space of all continuous functions is infinite because $1, x, x^2, \ldots$ is a set of linear independent functions.
- The dimension of the space of all polynomial functions is infinite because $1, x, x^2, \ldots$ is a basis.
- $\operatorname{Dim} V \oplus W = \operatorname{Dim} V + \operatorname{Dim} W$

**(3.24) Theorem.**
*Two finite dimensional vector spaces over $\mathbb{K}$ are isomorphic if and only if they have the same dimension.*

*Proof.* If $V$ and $W$ have the same dimension $n$ we can find two bases $\{v_1, \ldots, v_n\} \subset V$ and $\{w_1, \ldots, w_n\}$. The map

$$\phi : V \to W : \lambda_1 v_1 + \cdots + \lambda_n v_n \longmapsto \lambda_1 w_1 + \cdots + \lambda_n w_n$$

is welldefined because every element in $V$ has a unique presentation as a linear combination of the $v_i$. One can also check that it is linear. Finally, it is bijective because its inverse is the map

$$\phi^{-1} : V \to W : \lambda_1 w_1 + \cdots + \lambda_n w_n \longmapsto \lambda_1 v_1 + \cdots + \lambda_n v_n$$

Conversely if $\phi : V \to W$ is an isomorphism then the image of a basis is again a basis. The original basis spans $V$ and because $\phi$ is surjective the image of the basis spans $W$. The image of the basis is linearly independent because if there were a linear combination of the new basis that is zero then its image under $\phi^{-1}$ would give a linear combination of the original basis that is zero. □

We have now completed the task of classifying all finite dimensional vector spaces up to isomorphism. Every finite dimensional vector space is isomorphic to $\mathbb{K}^n$ for some $n$.

**(3.25) Aside.**
What about infinite dimensional vector spaces? Using standard set theory with the axiom of choice, one can show that every vector space has a basis, so infinite dimensional vector space have an infinite basis. But this does not mean that all infinite dimensional vector spaces are isomorphic. Some have bigger bases then others.
The correct way to restate the last theorem to infinite dimensional vector spaces is that $V$ and $W$ are isomorphic if and only if there exists a bijection between their bases.
For instance the space of maps from $\mathbb{N}$ to the reals that are zero for large $N \in \mathbb{N}$

$$V = \{\phi : \mathbb{N} \to \mathbb{R} | \exists N > 0 : \forall x > N : \phi(x) = 0\}$$

is an infinite dimensional vector space with basis

$$\phi_i : \mathbb{N} \to \mathbb{R} : \phi_i(j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

This space is isomorphic to to the space of polynomials $\mathbb{R}[X]$ because there is a bijection between

$$\phi_i \longmapsto x^i.$$

The vector space $\mathbb{Z}_2^{\mathbb{N}}$ is not isomorphic to $\mathbb{Z}_2[X]$. Suppose $\psi : \mathbb{Z}_2[X] \to \mathbb{Z}^{\mathbb{N}}$ were a linear bijection then we have a map $\psi'$ from $\mathbb{N} \to \mathbb{Z}^{\mathbb{N}}$ by demanding that

$$\psi'(n) = u \iff \exists f \in \mathbb{Z}_2[X] : f(2) = n \text{ and } \psi(f) = u.$$

By $f(2)$ we mean the number that we calculate from $f$ by filling in 2 and working in $\mathbb{Z}$ not in $\mathbb{Z}_2$ (e.g if $f(X) = [\![1]\!] + [\![1]\!]X^2$ then $f(2) = 5$). Note that the assignment $f \longmapsto f(2)$ is a bijection and therefore $\psi'$ is also a bijection. However, the map $\psi'$ cannot be not surjective because the vector

$$w = (\psi'_0(0) + 1, \psi'_1(1) + 1, \psi'_2(2) + 1, \dots)$$

where $\psi'_j(i)$ means the $j^{th}$ entry of the vector $\psi(i)$. Indeed $w \neq \psi'(j)$ because $w_j \neq \psi'_j(j)$. This argument is called Cantor's diagonalizations argument So we can conclude that not all infinite dimensional vector spaces over the same field are isomorphic.

# Chapter 4

# Maps and matrices

In this chapter, we shall assume that the vector space $V$ and $W$ are finite dimensional.

**(4.1) Definition.**
Suppose that $V$ and $W$ are finite dimensional vector spaces over $k$ with bases $\{v_1, v_2, \ldots, v_n\}$, $\{w_1, w_2, \ldots, w_m\}$, and suppose that $f$ is a linear map from $V$ to $W$. Let $A$ be the $m \times n$ matrix $A$ over $k$ whose $i$-th column

$$a^{(i)} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ \vdots \\ a_{mi} \end{pmatrix}$$

is defined by

$$f(v_i) = a_{1i}w_1 + a_{2i}w_2 + \ldots + a_{mi}w_m.$$

Then $A$ is called the matrix of $f$ with respect to the basis $\{v_1, \ldots, v_n\}$, $\{w_1, \ldots, w_m\}$.

We'll illustrate this with a simple example, and do further examples in a little while when we've understood the point of this definition.

**(4.2) Example.**
Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by the rule

$$f(a, b, c) = (a + 2b, c - a)$$

We can compute the matrix of $f$ with respect to the standard bases $e_1, e_2, e_3$ of $\mathbb{R}^3$ and $e_1, e_2$ of $\mathbb{R}^2$.

Since

$$\begin{aligned} f(e_1) &= f(1,0,0) = (1,-1) \\ f(e_2) &= f(0,1,0) = (2,0) \\ f(e_3) &= f(0,0,1) = (0,1) \end{aligned}$$

we see that the matrix is

$$A = \begin{pmatrix} 1 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

**(4.3) Proposition.**
*Suppose that $A$ is the matrix of a linear mapping $f : V \to W$ between finite dimensional vector spaces with bases $\{v_1, v_2, \ldots, v_n\}$, $\{w_1, w_2, \ldots, w_m\}$. Suppose that $x$ and $y$ are the column vectors of coefficients of a vector $\sum_{i=1}^{n} x_i v_i$ of $V$ and its image under $f$,*

$$\sum_{i=1}^{m} y_i w_i = f\left(\sum_{i=1}^{n} x_i v_i\right),$$

*that is*

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

*Then $x$ and $y$ are related by the equation*

$$Ax = y.$$

*Proof.*

$$\begin{aligned} \sum_{i=1}^{m} y_i w_i = f(\sum_{j=1}^{n} x_j v_j) &= f(x_1 v_1 + x_2 v_2 + \ldots + x_n v_n) \\ &= x_1 f(v_1) + x_2 f(v_2) + \ldots + x_n f(v_n) \\ &= (x_1 a_{11} + x_2 a_{12} + \ldots + x_n a_{1n}) w_1 \\ &+ (x_1 a_{21} + x_2 a_{22} + \ldots + x_n a_{2n}) w_2 \\ &+ \ldots\ldots + \ldots\ldots + \ldots\ldots \\ &+ \ldots\ldots + \ldots\ldots + \ldots\ldots \\ &+ (x_1 a_{m1} + x_2 a_{m2} + \ldots + x_n a_{mn}) w_m \end{aligned}$$

Comparing the coefficients of $w_1, w_2, \ldots$ in the left and right hand side of this equation, we see that

$$\begin{aligned} y_1 &= x_1a_{11} + x_2a_{12} + x_3a_{13} + \ldots + \ldots + x_na_{1n} \\ y_2 &= x_1a_{21} + x_2a_{22} + x_3a_{23} + \ldots + \ldots + x_na_{2n} \end{aligned}$$

etc. which we can write as a vector equation,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ \vdots \\ a_{m2} \end{pmatrix} + \ldots + \ldots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ \vdots \\ a_{mn} \end{pmatrix} = x_1a^{(1)} + x_2a^{(2)} + \ldots + x_na^{(n)}$$

and we have already observed, in subsection **??** that the right hand side of this is the matrix product $Ax$. □

**(4.4) Examples.** 1. Suppose that $V$ is the space of all polynomials over $k$ of degree at most $n$ and $W$ is the space of all polynomials over $k$ of degree at most $n+1$. Let $f : V \to W$ be the mapping defined by $f(p(x)) = xp(x)$.

We shall write down the matrix with respect to the bases $\{1, x, x^2, \ldots, x^n\}$ for $V$ and $\{1, x, x^2, x^{n+1}\}$ for $W$,

We can verify that matrix multiplication gives the correct image for a polynomial

$$a_0 + a_1x + a_2x^2 + \ldots a_nx^n$$

2. Let $V$ and $W$ be as above, and let $f$ be the mapping from $V$ to $W$ defined by $f(p(x)) = (x+2)p(x)$.

   We can write down the matrix with respect to the bases $\{1, x, x^2, \ldots, x^n\}$ for $V$ and $\{1, x, x^2, x^{n+1}\}$ for $W$,

3. Let $V$ be as above, and let $f$ be the mapping from $V$ to $V$ defined by $f(p(x)) = p(x+1)$. We'd like to write down the matrix for $f$ with respect to $\{1, x, x^2, \ldots x^n\}$. This is a little harder, so we'll restrict to the case $n = 5$. Once we've done this, it's clear what happens in general.

4. Let $V$ and $W$ both be equal to the space of polynomials of degree at most $n$, with basis $\{1, x, x^2, \ldots, x^n\}$, and let $f$ be the mapping from $V$ to $W$ defined by $f(p(x)) = p'(x)$ (the derivative of $p(x)$. Then $f(x^j)) = jx^{j-1}$ for each $j \geq 0$, and so we can easily write down the matrix for $f$.

5. Let $f : V_1 \to V_2$ and $g : W_1 \to W_2$ be represented by the matrices $A$ and $B$ for certain bases. These bases give rise to bases for $V_1 \oplus W_1$ and $V_2 \oplus W_2$ and the map $f \oplus g$ is represented by the matrix

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

Where the $0$ stands for a block of zeros of the appropriate size. We denote this matrix by $A \oplus B$.

**(4.5) Proposition.**
*Let $U, V, W$ be vector spaces over $k$, with bases $\{u_1, u_2, \ldots, u_n\}$, $\{v_1, \ldots, v_m\}$, $\{w_1, \ldots, w_l\}$, and let $f : U \to V$, $g : V \to W$ be linear mappings represented with respect to the given bases by matrices $A$ and $B$. Then the composite map $g \circ f$ is represented by the matrix $BA$.*

*Proof.* We simply have to compute $(g \circ f)(u_j)$ and verify that its coordinates are the entries in column $j$ of $BA$.
Now

$$\begin{aligned}
(g \circ f)(u_j) &= g(f(u_j)) = g(\sum_{k=1}^{m} A_{kj} v_k) = \sum_{k=1}^{m} A_{kj} g(v_k) \\
&= \sum_{k=1}^{m} A_{kj} (\sum_{i=1}^{l} B_{ik} w_i) = \sum_{k=1}^{m} \sum_{i=1}^{l} A_{kj} B_{ik} w_i \\
&= \sum_{i=1}^{l} \sum_{k=1}^{m} A_{kj} B_{ik} w_i \quad \text{we can change the order of summation} \\
&= \sum_{i=1}^{l} \sum_{k=1}^{m} B_{ik} A_{kj} w_i \quad \text{we can change the order within each product} \\
&= \sum_{i=1}^{l} (BA)_{ij} w_i
\end{aligned}$$

□

# 1 Base change

**(4.6) Proposition.**
*Suppose that $V, W$ are vector spaces over $k$. Let $\{v_1, \ldots, v_n\}$ be a basis of $V$, and suppose that $w_1, \ldots, w_n$ are elements of $W$. Then there is a unique linear mapping $f : V \to W$ with $f(v_i) = w_i$ for each $i$.*

*Proof.* Define $f$ by $f(\sum_{i=1}^{n} \lambda_i v_i) = \sum_{i=1}^{n} \lambda_i v_i$. Then $f$ is well defined as a mapping from $V$ to $W$ since each element of $V$ can be uniquely represented as a sum $\sum_{i=1}^{n} \lambda_i v_i$.
To verify linearity notice that

$$\begin{aligned} f(\alpha \sum_{i=1}^{n} \lambda_i v_i) &= f(\sum_{i=1}^{n} \alpha\lambda_i v_i) \\ &= \sum(\alpha\lambda_i) w_i \\ &= \alpha \sum(\lambda_i w_i) \\ &= \alpha f(\sum \lambda_i v_i) \end{aligned}$$

and also that

$$\begin{aligned} f(\sum_{i=1}^{n} \lambda_i v_i + \sum_{i=1} \mu_i v_i) &= f(\sum(\lambda_i + \mu_i) v_i) \\ &= \sum(\lambda_i + \mu_i) w_i \\ &= \sum \lambda_i w_i + \sum \mu_i w_i \\ &= f(\sum \lambda_i v_i) + f(\sum \mu_i v_i) \end{aligned}$$

$f$ must be unique, since any linear mapping satisfying $f(v_i) = w_i$ has to satisfy

$$f(\sum \lambda_i v_i) = \sum \lambda_i f(v_i) = \sum \lambda_i w_i$$

□

**(4.7) Corollary. (Rather obvious?)**
*A linear mapping from a vector space $V$ to itself is the identity if it fixes a basis of $V$.*

**(4.8) Proposition.**
*Let $V$ be a vector space over $k$ with basis $\{v_1, v_2, \ldots v_n\}$, let $w_1, \ldots, w_n$ be elements of $W$, and let $f$ be the linear mapping $V \to W$ defined by $f(v_i) = w_i$. Then $f$ is*

- *surjective if and only if $\langle w_1, \ldots, w_n \rangle = W$,*
- *injective if and only if $w_1, \ldots, w_n$ are linearly independent,*
- *bijective (=invertible) if and only if $w_1, \ldots, w_n$ is a basis for $W$,*

*Proof.* The first assertion follows from the fact that

$$\mathrm{Im}\, f = \{f(v)|v \in V\} = \{f(\sum_i a_i v_i)|a_i \in \mathbb{K}\} = \{\sum_i a_i w_i|a_i \in \mathbb{K}\} = \langle w_1, \ldots, w_n\rangle$$

The second holds because $\sum_i a_i w_i = 0$ if and only if $f(\sum_i a_i v_i) = 0$ i.e. $\sum_i a_i v_i \in \mathrm{Ker}\, f$. The third statement is a combination of the first two. □

**(4.9) Corollary.**
*Suppose that $V$ is a vector space over $k$ with basis $\{v_1, \ldots, v_n\}$. Then where $A$ is an $n \times n$ matrix over $k$, $\{\sum_{i=1}^n A_{ij} v_i\}$ is a basis of $V$ if and only if $A$ is invertible.*

*Proof.* $A$ is the matrix with respect to $\{v_i\}$ of the linear mapping $f$ which maps $v_j$ to $\sum A_{ij} v_i$ for each $j$. □

**(4.10) Definition.**
Where $V$ is a vector space with bases $\{v_i\}$ and $\{v_i'\}$ the matrix $A$ defined by $v_j' = \sum_{i=1}^n A_{ij} v_i$ is called the matrix for the base change from $\{v_i\}$ to $\{v_i'\}$. Clearly $A^{-1}$ is the matrix for the base change from $\{v_i'\}$ to $\{v_i\}$.

Now if $v$ is a vector in $V$ with

$$v = \sum_{i=1}^n x_i v_i = \sum_{j=1}^n y_j v_j'$$

then

$$\begin{aligned} v &= \sum_{j=1}^n y_j v_j' \\ &= \sum_{j=1}^n y_j \sum_{i=1}^n A_{ij} v_i \\ &= \sum_{i=1}^n (\sum_{j=1}^n A_{ij} y_j) v_i \\ &= \sum_{i=1}^n (A\mathbf{y})_i v_i \end{aligned}$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$$

That is, the vector $\mathbf{x}$ of coefficients of a vector $v$ with respect to the basis $\{v_i\}$ is related to the vector $\mathbf{y}$ of coefficients of $v$ with respect to the basis $\{v'_i\}$ by th matrix equation

$$x = Ay$$

**(4.11) Theorem.**
*Suppose that $V$ and $W$ are vector spaces over $k$, that $\{v_1, \ldots, v_n\}$ and $\{v'1_1, \ldots, v'_n\}$ are bases for $V$, and that $\{w_1, \ldots, w_m\}$ and $\{w'_1, \ldots, w'_m\}$ are bases for $W$.*
*Let $f$ be a linear mapping from $V$ to $W$, and suppose that $f$ has matrix $T$ with respect to the bases $\{v_i\}$ and $\{w_i\}$ and matrix $T'$ with respect to the bases $\{v'_i\}$ and $\{w'_i\}$. Suppose also that $P$ and $Q$ are the matrices for the base changes from $\{v_i\}$ to $\{v'_i\}$ and from $\{w_i\}$ to $\{w'_i\}$. Then*

$$T' = Q^{-1}TP$$

*Proof.* For each $j$,

$$\begin{aligned} f(v'_l) = f(\sum_{k=1}^{n} P_{kl} v_k) = \sum_{k=1}^{n} P_{kl} f(v_k) & \\ &= \sum_{k=1}^{n} (P_{kl} \sum_{j=1}^{m} T_{jk} w_j) \\ &= \sum_{k=1}^{n} (P_{kl} \sum_{j=1}^{m} (T_{jk} \sum_{i=1}^{m} (Q^{-1})_{ij} w'_i \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{n} (Q^{-1})_{ij} T_{jk} P_{kl} w'_i \\ &= \sum_{i=1}^{m} (Q^{-1}TP)_{il} w'_i \end{aligned}$$

Thus $Q^{-1}TP$ is the matrix representing $f$ with respect to the bases $\{v'_i\}$ and $\{w'_i\}$. Hence (since $T'$ is defined to be that matrix)

$$T' = Q^{-1}TP$$

□

**(4.12) Corollary.**
*Suppose that $T$ and $T'$ are two $m \times n$ matrices, and that for some invertible matrices $A, B$, $T' = ATB$. Then $T$ and $T'$ represent the same linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ with respect to different bases.*

*Proof.* In the above, define $\{v_i\} = \{e_i\}$, $\{w_i\} = \{e_i\}$, and let $v'_j = \sum B_{ij}v_i$, $w'_j = \sum A^{-1}_{ij}w_i$. □

We've seen now that matrices are a nice way to present linear maps. Given two bases: one for $V$ and one for $W$ we have a bijection

$$\mathrm{Hom}(V, W) \to \mathrm{Mat}_{\dim W \times \dim V}.$$

This bijection is even a linear map itself. As we already said this bijection can only be made using the bases and if one takes other bases we get another bijection.
This gives us freedom to choose different bases such that we can try to find a nice way to represent a given map as a matrix. This leads to the following questions.

**(4.13) Question.**
*Given a linear map $f$ is there some kind of standard form we can bring the matrix that represents $f$, in*

1. *using base changes in both the source and the target?*
2. *using base changes only in the target (or only in the source)?*
3. *using base changes in the target if source and target coincide?*

We can formulate them in matrix-only language:

**(4.14) Question.**
*Given an $n \times m$-matix $T$ is there some kind of standard form we can bring it in*

1. *using transformations of the form $T \longmapsto ATB$?*
2. *using transformations of the form $T \longmapsto AT$ (or $T \longmapsto TB$)?*
3. *using transformations of the form $T \longmapsto A^{-1}TA$ if $n = m$?*

*Here $A$ and $B$ are invertible matrices.*

In the next 3 chapters we will study each of these 3 questions more closely.

# Chapter 5

# Rank

## 0.1 Kernel and Image

Linear maps also enable us to define subspaces.

**(5.1) Definition.**
Given a linear map $\phi : V \to W$ we define the image of $\phi$ as the set of all elements of $W$ reached by the map and the kernel as all elements that are mapped to zero.

- $\operatorname{Im}\phi := \{w \in W | \exists v \in V : w = \phi(v)\}$
- $\operatorname{Ker}\phi := \{v \in V | \phi(v) = 0\}$

**(5.2) Lemma.**
*$\operatorname{Im}\phi$ is a subspace of $W$ and $\operatorname{Ker}\phi$ is a subspace of $V$.*

*Proof.* If $w_1 = \phi(v_1)$ and $w_2 = \phi(v_2)$ then $\lambda w_1 + \mu w_2 = \phi(\lambda v_1 + \mu v_2)$ so (*) holds for $\operatorname{Im}\phi$.
If $\phi(v_1) = 0$ and $\phi(v_2) = 0$ then $\phi(\lambda v_1 + \mu v_2) = \lambda 0 + \mu 0 = 0$ so (*) holds for $\operatorname{Ker}\phi$. □

The image and the kernel can be used to translate injectivity and surjectivity to linear algebra.

**(5.3) Lemma.**
*A map $\phi : V \to W$ is*

- *surjective if and only if* $\operatorname{Im}\phi = W$,
- *injective if and only if* $\operatorname{Ker}\phi = \{0\}$.

*Proof.* The first statement follows directly from the definition of surjectivity. If $\phi$ is injective there can only be one element in $V$ that is mapped to the zero. By the linearity, this unique element must be the zero $0$. If $\phi$ is not injective then we can find $v_1, v_2 \in V$ with $\phi(v_1) = \phi(v_2)$. The difference $v_1 - v_2$ sits in the kernel. □

## 0.2 Quotient spaces

Given a vector space $V$ and a subspace $U \subset V$ we can construct an equivalence relation on $V$

$$v \cong_U w \iff v - w \in U$$

This is indeed an equivalence relation: reflexivity follows from the fact that $0 \in U$, symmetry from $u \in U \implies -u \in U$ and transitivity from $u_1, u_2 \in U \implies u_1 + u_2 \in U$.
This equivalence relation is compatible with the structure of the the vector space

- if $v_1 \cong_U v_2$ then $\lambda v_1 \cong_U \lambda v_2$,
- if $v_1 \cong_U v_2$ and $w_1 \cong_U w_2$ then $v_1 + w_1 \cong_U v_2 + w_2$

Therefore we can turn the quotient set $V/U := V/\cong_U$ into a vector space by setting $\lambda[\![v]\!] = [\![\lambda v]\!]$ and $[\![v]\!] + [\![w]\!] = [\![u + w]\!]$. This definition also implies that the quotient map

$$\pi : V \to V/U : v \longmapsto [\![v]\!]$$

is a linear surjection.

**(5.4) Theorem.**
*If $U$ is a subspace of $V$ and $V$ is finite dimensional, then $V/U$ is finite dimensional and*

$$\operatorname{Dim} V/U = \operatorname{Dim} V - \operatorname{Dim} U$$

*Proof.* Choose a basis $u_1, \ldots, u_k$ for $U$ and extend it to a basis for $V$: $\{u_1, \ldots u_k, v_1, \ldots v_l\}$ The dimension of $U$ is $k$ and of $V$ is $k+l$. We now show that $\{[\![v_1]\!], \ldots, [\![v_l]\!]\}$ is a basis for $V/U$. It is clear that $\langle [\![v_1]\!], \ldots, [\![v_l]\!] \rangle$ is $V/U$ because

$$[\![\lambda_1 u_1 + \cdots + \lambda_k u_k + \mu_1 v_1 + \cdots + \mu_l v_l]\!] = [\![\mu_1 v_1 + \cdots + \mu_l v_l]\!] = \mu_1 [\![v_1]\!] + \cdots + \mu_l [\![v_l]\!].$$

Furthermore the $\{[\![v_1]\!], \ldots, [\![v_l]\!]\}$ are linearly independent because if $\mu_1 [\![v_1]\!] + \cdots + \mu_l [\![v_l]\!] = 0$ then $\mu_1 v_1 + \cdots + \mu_l v_l \in U$ but because $\{u_1, \ldots u_k, v_1, \ldots v_l\}$ is a basis of $V$ the $\mu_i$ must be zero. □

**(5.5) Examples.**

Let $V$ be the subset of $\mathcal{P}(S)$ containing all sets with an even number of elements. The quotient $\mathcal{P}(S)/V$ contains two elements the set of all even subsets and the set of all odd subsets. Hence the quotient is isomorphic to $\mathbb{Z}_2^1$.

Let $V$ be the subspace of $\mathbb{C}^0(\mathbb{R})$ containing all functions for which $f(0) = 0$. The equivalence classes of $\cong_V$ are $[\![g]\!] = \{f \in C^0(\mathbb{R}) | f(0) = g(0)\}$ and we can parametrize them by $g(0)$. The quotient space is therefore isomorphic with $\mathbb{R}$ by the linear map $\mathbb{C}^0(\mathbb{R})/V \to \mathbb{R} : [\![g]\!] \longmapsto g(0)$.

Let $V$ be the subspace of $\mathbb{R}^3$ spanned by $(1, 1, 1)$. The equivalence classes of $\cong_V$ contain a unique element from $U = \{(0, a, b) | a, b \in \mathbb{R}\}$ so we have a bijective map

$$\phi : U \to \mathbb{R}^3/V : (0, a, b) \longmapsto [\![(0, a, b)]\!]$$

**(5.6) Theorem.**
*For any linear map $\phi : V \to W$ we have that*

$$\operatorname{Im} \phi = V/\operatorname{Ker} \phi.$$

*Proof.* We have a map

$$\bar{\phi} : V/\operatorname{Ker} \phi \to \operatorname{Im} \phi : [\![v]\!] \to \phi(v).$$

This map is well defined and injective because $[\![v]\!] = [\![w]\!] \iff \phi(v) = \phi(w)$. The map is surjective because the target is $\operatorname{Im} \phi$. □

**(5.7) Corollary.**
*Every linear map $\phi : V \to W$ can be written as a composition*

$$\iota \circ \bar{\phi} \circ \pi$$

*where*

- $\pi : V \to V/\operatorname{Ker} \phi : v \longmapsto [\![v]\!]$ *is a surjection,*
- $\bar{\phi} : V/\operatorname{Ker} \phi \to \operatorname{Im} \phi : [\![v]\!] \longmapsto \phi(v)$ *is a bijection,*
- $\iota : \operatorname{Im} \phi \to W : [\![w]\!] \longmapsto w$ *is an injection,*

**(5.8) Definition.**
For any linear map we will call $\operatorname{Dim} \operatorname{Im} \phi = \dim V - \operatorname{Dim} \operatorname{Ker} \phi$ the rank of $\phi$. If $A$ is a matrix then its rank, is the rank of the corresponding linear map $f_A : \mathbb{R}^m \to R^n$.

We will see now that the rank plays an important role in the classification of linear maps.

# 1 Base Change and Rank

**(5.9) Definition.**
We say two $m \times n$ matrices $T$ and $T'$ are basechange equivalent if and only if there exist invertible matrices such that $T' = ATB$ We denote this by $\cong_B$.

**(5.10) Theorem.**
*$\cong_B$ is an equivalence relation on* $\mathrm{Mat}_{m\times n}(\mathbb{K})$.

*Proof.* $T \cong_B T$ because $T = 1T1$. If $T' = ATB$ then $T = A^{-1}T'B^{-1}$ so $\cong_B$ is symmetric. Finally, transitivity follows because if $T' = ATB$ and $T'' = CT'D$ then $T'' = (CA)T(BD)$. □

The question we can ask now is how do the equivalence classess look like.

**(5.11) Theorem.**
*Let $T$ be any $m \times n$ matrix. Then there is an invertible $m \times m$ matrix $A$ and an invertible $n \times n$ matrix $B$ such that*

$$ATB = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0m-r, n-r \end{pmatrix}$$

*Proof.* $T$ represents a linear mapping $f$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, with respect to the bases $\{e_i\}_{i=1}^n$ and $\{e_i\}_{i=1}^m$. Construct a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that $\{v_{r+1}, \ldots, v_n\}$ is a basis for $ker(f)$. Then choose a basis $\{w_1, \ldots, w_m\}$ of $\mathbb{R}^m$ such that $w_i = f(v_i)$ for $i = 1, \ldots, r$. This is indeed possible because the $f(v_i)$ for $i = 1, \ldots, r$ form a linearly independent set ($\sum_{i=1}^r a_i f(v_i) = 0$ implies $\sum_{i=1}^r a_i v_i \in \mathrm{Ker}\, f$ contradicting that $v_i, i > r$ is a basis for $\mathrm{Ker}\, f$. With respect to the bases $\{v_i\}$ and $\{w_i\}$, $f$ is represented by the matrix

$$T' = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0m-r, n-r \end{pmatrix}.$$

Now where $A^{-1}$ is the matrix for the base change of $\mathbb{R}^m$ from $\{e_i\}$ to $\{w_i\}$ and $B$ is the matrix for the base change of $\mathbb{R}^n$ from $\{e_i\}$ to $\{v_i\}$, by the previous result we have

$$ATB = T'$$

□

**(5.12) Definition.**
We will call $r$ the rank of the matrix $T$.

**(5.13) Corollary.**
*If $f : V \to W$ is a map and $A$ a matrix representing it then* $\operatorname{Rank} A = \dim \operatorname{Im} f = \dim V - \dim \operatorname{Ker} f$.
$\operatorname{Rank} A$ *is also the dimension of the span of the columns (which is equal to* $\dim \operatorname{Im} f$*) and because* $\operatorname{Rank} A = \operatorname{Rank} A^\top$*,* $\operatorname{Rank} A$ *is also the dimension of span of the rows.*

**(5.14) Theorem.**
*The equivalence classes of* $\mathrm{Mat}_{m\times n}, \cong_B$ *are parametrized by the rank:* $\mathrm{Mat}_{m\times n}/\cong_B$ *has* $\min(m,n)+1$ *elements.*

$$\mathrm{Mat}_{m\times n}/\cong_B = \{\{A \in \mathrm{Mat}_{m\times n} | \operatorname{Rank} A = r\} | 0 \leq r \leq n, m\}$$

*Proof.* We have already seen that every matrix sits in the class of some

$$T' = \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0m-r, n-r \end{pmatrix}.$$

Two such matrices with different $r$ cannot sit in the same class because base changes do not change the rank of a matrix.
Finally $r \leq n, m$ because otherwise we cannot fit $\mathbb{I}_n$ inside an $n \times m$-matrix.
Hence $r = 0, 1, \ldots, \min(n, m)$. □

Although we have a nice classification of linear maps by means of its rank, we still do not have an easy way to calculate the rank, image and kernel of a given map or matrix. In the next chapter we will see such a method.

# Chapter 6

# Row echelon matrices

## 1 Base change on one side

In some problems it is interesting to fix a basis in the target or the source of the map and try to find a basis in the source or the target such that the matrix that represents the map has a nice form.

Suppose we want to describe the image of the map $f : V \to W$ in terms of a basis of $W$, then we still have a choice for the basis in the source. Similarly if we want to describe the kernel of the map in terms of a given basis in $V$ we still can vary the basis in $W$.

Just like in the previous chapter we formulate this problem in the language of equivalence relations.

**(6.1) Definition.**
Two $m \times n$-matrices $T, T'$ are row or left equivalent if there is an invertible matrix $A$ such that $T' = AT$. We denote this by $T' \cong_L T$. Similiarly we can define a notion of column or right equivalence.

**(6.2) Theorem.**
*$\cong_L$ and $\cong_R$ are equivalence relations on $\mathrm{Mat}_{m\times n}(\mathbb{K})$.*

*Proof.* Analogous to **??** □

**(6.3) Theorem.**
*Let $f : V \to W$ and $g : V \to W$ be two linear maps and let $A$ and $B$ be their matrix representations according to two fixed bases. If $A \cong_L B$ then $f$ and $g$ have the same kernel, if $A \cong_R B$ then $f$ and $g$ have the same image.*

## 2 The Row Echelon Form

The problem is now to describe the equivalence classes of $\cong_L$.

**(6.4) Definition. Matrix in row echelon form**
We say that a matrix $A$ is a row echelon matrix if the first non-zero entry in each row is further right than the corresponding entry in the row above it. If additionally the first non-zero entry of each row is a $1$ and this $1$ is the only nonzero entry in its column then we speak of a reduced row echelon matrix.

If the coefficient matrix of a system of simultaneous linear equations is in echelon form it is relatively easy to solve the system.
Example:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 15 \\ x_2 + 2x_3 + 2x_4 + 2x_5 &= 26 \\ x_3 + x_4 + x_5 &= 12 \\ x_4 - x_5 &= -1 \\ x_5 &= 5 \end{aligned}$$

We solve from the bottom up, i.e. for $x_5$, then $x_4$ etc.:-

## 3 Row operations

We'll use the techniques of Gaussian elimination to transform a given matrix or set of vectors in $\mathbb{R}^n$ into one in row echelon form; this allows us to solve a number of related problems. Gaussian elimination is simply a sequence of carefully chosen row operations.

**(6.5) Definition.**
When $A$ is an $m \times n$ matrix we define an elementary row operation on $A$ to be any one of the following types of operations on the rows of $A$:-

1 one row of $A$ is replaced by a non-zero **multiple** of itself

2 two rows of $A$ are **swapped**

3 one row of $A$ is replaced by the **sum** of itself and a scalar multiple of another row

The row operations can be thought of as multiplying $A$ on the left with special matrices.

1 Rescaling the $i^{th}$ row with factor $\lambda$ is the same as transforming $A$ into $RA$ where $R$ is the identity matrix except on the $i, i^{th}$ entry where $R_{ii} = \lambda$.

2 Swapping the $i^{th}$ row with the $j^{th}$ is the same as transforming $A$ into $RA$ where $R$ is the identity matrix except for $R_{ij} = R_{ji} = 1$ and $R_{ii} = R_{jj} = 0$.

3 Adding $\lambda$ times the $j^{th}$ row to the $i^{th}$ is the same as transforming $A$ into $RA$ where $R$ is the identity matrix except for $R_{ij} = \lambda$.

All these matrices are invertible, so if $A'$ is obtained from $A$ after applying row operations then $A' \cong_L A$.

# 4 Gaussian Elimination

**(6.6) Definition. pivot**
We define a pivot in a matrix $A$ to be the first non-zero entry in some row. We say that a selected pivot is leftmost from row $i$ in $A$ if there is no pivot in $A$ to the left of the selected pivot in or below row $i$.

**(6.7) Algorithm. Standard Gaussian elimination**

**Input:** An $m \times n$ matrix $A$.
For $i = 1$ to $m - 1$ do:

**if** there is no pivot in row $i$ or below, halt,

**otherwise** :

**swap** rows as necessary to ensure that there's a small leftmost pivot in row $i$,

**divide** row $i$ by the pivot.

**zero in the pivot column** that is, where the pivot in row $i$ is in column $j$, subtract an appropriate multiple of row $i$ from each row below and above it so that there are zeros in column $j$ below and above row $i$.

**Output:** An $m \times n$ matrix $A'$ in reduced row echelon form.
NB: If one only needs the row echelon form one can omit the step divide and change the step [zero in the pivot column] to [zero below the pivot].

**(6.8) Examples.**

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 1 & 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{matrix} R2 \to R1 - 5R1 \\ R3 \to R3 - 9R1 \\ R4 \to R4 - R1 \end{matrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -16 & -24 \\ 0 & -1 & -2 & -3 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \\ 1 & 0 & 1 & 0 \\ 2 & 3 & 2 & 2 \end{pmatrix} \longrightarrow$$

**(6.9) Theorem.**
*If two matrices in reduced row echelon form are row equivalent then they are the same*

*Proof.* Let $T$ be a reduced row echelon matrix and $A$ an invertible matrix such that $AT$ is also a reduced row echelon matrix. We show that $AT = T$. We do this by induction on the nonzero rows of $T$. If al rows of $T$ are zero then $T = 0$ so $AT = T = 0$.
Suppose that the propostion holds for all matrices with at most $k$ nonzero rows and let $T$ be a matrix with $k + 1$ nonzero rows.
Let $p$ be the position of the first nonzero column of $T$ and hence also of $AT$. If $A_{i1} \neq 0$ for some $i > 0$ then $AT$ cannot be in RREF because $(AT)_{ip} = A_{ip} \neq 0$. Therefore $A_{i1} = 0$ for $i > 0$ and the rows of $AT$ with index bigger than $1$ are linear combinations of the rows of $A$ with index bigger than $1$.
Let $\hat{T}$ and $\widehat{(AT)}$ are the denote the matrices obtained by deleting the first row of $T$ and $AT$, while $\hat{A}$ stands for the matrix constructed by deleting the first row and column of $A$. By the previous discussion we have that

$$\widehat{(AT)} = \hat{A}\hat{T}$$

If one removes the upper row of a row reduced echelon matrix, it is still a row reduced echelon matrix. so by the induction hypothesis $\widehat{(AT)} = \hat{A}\hat{T}$.
We only need to show that the first row of $AT$ is equal to the first row of $T$. If this were not the case $A_{1j} > j$ for some $j > 1$. One can calculate that $(AT)_{1p_j} = A_{1p_j}$ where $p_j$ is the index of the pivot of the $j^{th}$ row, but $T_{1p_j}$ is zero because $T$ is a row reduced echelon matrix. □

This theorem shows that every equivalence class contains a unique reduced row echelon matrix. De elements in the quotient $\mathrm{Mat}_{m\times n}/\cong_L$ can be parametrized by the reduced row echelon matrices.

# 5 Solving Problems using the row echelon form

Gaussian elimination can be used to solve many problems in linear algebra

**(6.10) Problem.**
*Determining the rank of a given matrix.*

**(6.11) Solution.**
Bring the matrix in reduced row echelon form. The rank is the number of nonzero rows.

**(6.12) Problem.**
*Determining the dimension of the span of a set of vectors.*

**(6.13) Solution.**
Consider the matrix with rows the given vectors. Bring the matrix in reduced row echelon form. The dimension of the span is the number of nonzero rows.

**(6.14) Problem.**
*Determining whether a given vector $v$ is in the span of a given set of vectors.*

**(6.15) Solution.**
Consider the matrix with rows the given set of vectors. Bring the matrix in reduced row echelon form. For every nonzero row in $A$ substract a multiple of from $v$ to make the coefficient of the corresponding pivot $0$. If after these operation $v$ becomes zero then $v$ is in the span else $v$ is not.

**(6.16) Problem.**
*Determining a set of vectors that span the kernel of a given matrix $A$.*

**(6.17) Solution.**
Bring the matrix in reduced row echelon form. Let $S \subset \{1, \ldots, n\}$ be the set of column numbers for which there is no pivot. For every $i \in S$ we define the vector $w$ that has a $1$ on position $i$ and for every $j \in \{1, \ldots, n\} \setminus S$ we put $-A_{lj}$ where $l$ is the row number of the pivot.

**(6.18) Problem.**
*Finding a matrix such that the span of a set of vectors is its kernel.*

**(6.19) Solution.**
Consider the matrix with rows the given vectors and transpose it. Bring the matrix in reduced row echelon form. Determine the kernel of this map. The matrix with as row vectors a basis for the kernel, is the matrix we are looking for.

**(6.20) Problem.**
*Finding the inverse of an $n \times n$-matrix $A$.*

**(6.21) Solution.**
Construct the $2n \times n$-matrix $[A\ \mathbb{I}_n]$ and bring it in reduced row echelon form. If the matrix is invertible, then the reduced row echelon form looks like $[\mathbb{I}_n A^{-1}]$. If the reduced row echelon form does not start with $\mathbb{I}_n$ the matrix is not invertible.

# Chapter 7

# Conjugation and Eigenvectors

So far we classified maps under base change in their source and target together or separately. However we did not consider yet maps for which their source and target coincide.
In this case we get a new type of equivalence relation.

**(7.1) Definition.**
Two $n \times n$-matrices $T$ and $T'$ are conjugated if there is an invertible matrix $A$ such that $T' = ATA^{-1}$. We denote this by $T \cong_C T'$

**(7.2) Theorem.**
*$\cong_C$ is an equivalence relation.*

Again the task is to classify the equivalence classes of matrices under conjugation. In general this problem is more complicated then the previous two. Unlike the former the answer depends on the structure of the field over which we are working.
The solution will be the simplest when the field is algebraicly closed for instance if we are working over $\mathbb{C}$.

## 1 Eigenvectors

**(7.3) Definition.**
Suppose that $V$ is a vector space over a field $\mathbb{K}$, and that $f : V \to V$ is a linear mapping. Then a non-zero vector $v \in V$ is called an eigenvector of $f$ if

$$f(v) \in \langle v \rangle$$

In that case, there is some $\lambda$ in $k$ with

$$f(v) = \lambda v$$

$\lambda$ is called the eigenvalue of $f$ associated with $v$.
Where $A$ is a matrix which represents $f$ with respect to some basis of $V$, we also say that $v$ is an eigenvector of $A$, and $\lambda$ is an eigenvalue of $A$.

**(7.4) Examples.** 1. Where $V = \mathbb{R}[x]$, the set of polynomials over $\mathbb{R}$, and $D : V \to V$ is the linear mapping defined by

$$D(p(x)) = p'(x)$$

, then every constant polynomial is an eigenvector, with eigenvalue 0. For if $p(x) = c$, $D(p(x)) = 0 = 0p(x)$.

If we are looking at $D$ defined on $C^1(\mathbb{R})$ we see that $f(x) = e^{ax}$ is an eigenvector with eigenvalue $a$: $Df(x) = af(x)$.

2. Where $V = \mathbb{R}^3$ and $f$ is the linear mapping represented with respect to the standard basis $e_1, e_2, e_3$ by the matrix

$$A = \begin{pmatrix} 3 & 2 & 0 \\ 0 & 4 & 0 \\ 0 & 6 & 3 \end{pmatrix}$$

then $e_1$, $e_3$, and in fact any vector of the form $ae_1+be_3$ are eigenvectors with eigenvalue 3.

3. Where $V = \mathbb{R}^3$ and $f$ is the linear mapping represented with respect to the standard basis $e_1, e_2, e_3$ by the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 0 & 0 & 3 \end{pmatrix}$$

then $e_1 + e_2$ is an eigenvector with eigenvalue 1.

How do we find the eigenvectors and eigenvalues of a linear mapping $f$?

**(7.5) Proposition.**
*Suppose that $V$ is a vector space over $k$ and that $f : V \to V$ is a linear mapping, represented with respect to a basis $\{v_1, \ldots, v_n\}$ of $V$ by an $n \times n$ matrix $A$. Then*

*$v = \sum x_i v_i$ is an eigenvector of $V$ with eigenvalue $\lambda$ if and only if the column vector*

$$x = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{pmatrix}$$

*is a non-zero solution to the matrix equation*

$$(A - \lambda \mathbb{I}_n)x = 0$$

Hence we find the eigenvalues of $f$ by finding all possible $\lambda$ such that the matrix $A - \lambda\mathbb{I}_n$ is singular, that is, we find all $\lambda$ such that $A - \lambda\mathbb{I}_n$ has determinant 0. The equation $\det(A - \lambda\mathbb{I}_n) = 0$ has degree $n$ and so at most $n$ solutions, so there can be at most $n$ distinct eigenvalues.
Where $\mathbb{K} = \mathbb{C}$, every polynomial over $\mathbb{C}$ has $n$ roots, and hence $A$ has $n$ distinct eigenvalues iff no eigenvalue is repeated.
Then for a given eigenvalue $\lambda$ we find the coefficients of a corresponding eigenvector $v$ (in terms of the chosen basis) by solving the matrix equation

$$(A - \lambda \mathbb{I}_n)x = 0$$

Each eigenvalue has many eigenvectors. For instance, any multiple of an eigenvector is an eigenvector. The eigenvectors of an eigenvalue actually form a subspace of the whole space, called an eigenspace, which may have dimension more than 1.
If $A$ is diagonal, then the eigenvalues of $A$ are very easy to find they are simply the diagonal entries of $A$, and the standard basis vectors of $\mathbb{R}^n$ are all eigenvectors.
If all the entries of $A$ on one side of the diagonal are 0 then the eigenvalues of $A$ are again the diagonal entries, but the standard basis vectors are not usually eigenvectors. (If all the entries on the lower side of the diaginal are zero, we say that $A$ is upper-triangular. If all the entries on the upper side of the diagonal are zero, we say that $A$ is lower-triangular.)

**(7.6) Definition.**
Where $A$ is an $n \times n$ matrix and where $B^{-1}AB$ is diagonal for some $n \times n$ matrix $B$, we say that $A$ is diagonalisable.

**(7.7) Proposition.**
*An $n \times n$ matrix $A$ is diagonalisable if and only if $A$ has $n$ linearly independent eigenvectors $v_1, v_2, \ldots v_n$.*

*Proof.* Let $f$ be the mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ represented with respect to the standard basis by $A$. If $A$ is diagonalisable, then for some $B$, $B^{-1}AB$ is diagonal. $B^{-1}AB$ represents $f$ with respect to the basis whose elements are the columns of $B$. Since $B^{-1}AB$ is diagonal, each of those vectors is mapped to a multiple itself, and hence is an eigenvector. So $A$ has $n$ linearly independent eigenvectors.
Conversely, if $A$ has $n$ linearly independent eigenvectors, $v_1, \ldots, v_n$ let $B$ be the matrix for the base change from the standard basis to $\{v_1, \ldots, v_n$. Then $B^{-1}AB$ is diagonal, with the eigenvalues of $A$ as its diagonal entries. □

Some $n \times n$ matrices are diagonalisable. Some are not. the following is an important result.

**(7.8) Theorem.**
*Suppose that $V$ is a vector space over $\mathbb{K}$ and that $f : V \to V$ is a linear mapping. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct eigenvalues of $f$, then where $v_1, v_2, \ldots, v_n$ are the associated eigenvectors, $\{v_1, \ldots, v_n\}$ are linearly independent vectors.*

NB. This result does not require $n = \dim V$, or even that $\lambda_1, \ldots, \lambda_n$ are all the eigenvalues of $f$.

*Proof.* The proof is by induction on $n$.
When $n = 1$ there is just one eigenvector, $\{v_1\}$, which is non-zero by definition, and so $\{v_1\}$ is a linearly independent set.
So suppose that the result is true for a set of less than $n$ distinct eigenvalues. Then the vectors $v_1, \ldots, v_{n-1}$, the eigenvectors for $\lambda_1, \ldots, \lambda_{n-1}$, are linearly independent.
So now suppose that the vectors $v_1, \ldots, v_n$ are linearly dependent. Then for some $\mu_1, \ldots, \mu_n$,

$$\mu_1 v_1 + \mu_2 v_2 + \ldots \mu_n v_n = 0$$

where not all $\mu_i$ are zero. In particular $\mu_n \neq 0$.
Applying $f$ to the above, we see that

$$\begin{aligned}
0 &= f(0) = f(\mu_1 v_1 + \mu_2 v_2 + \ldots \mu_n v_n) \\
&= \mu_1 f(v_1) + \mu_2 f(v_2) + \ldots \mu_n f(v_n) \\
&= \lambda_1\mu_1 v_1 + \lambda_2\mu_2 v_2 + \ldots + \lambda_n\mu_n v_n \\
\text{Hence } 0 &= \lambda_k(\mu_1 v_1 + \mu_2 v_2 + \ldots \mu_n v_n) - (\lambda_1\mu_1 v_1 + \lambda_2\mu_2 v_2 + \ldots + \lambda_n\mu_n v_n) \\
&= (\lambda_n - \lambda_1)\mu_1 v_1 + (\lambda_n - \lambda_2)\mu_2 v_2 + \ldots (\lambda_n - \lambda_n)\mu_n v_n \\
0 &= (\lambda_n - \lambda_1)\mu_1 v_1 + (\lambda_n - \lambda_2)\mu_2 v_2 + \ldots (\lambda_n - \lambda_{n-1})\mu_{n-1} v_{n-1}
\end{aligned}$$

Now, by the induction hypothesis, the vectors $v_1, v_2, \ldots, v_{n-1}$ are linearly independent. So

$$(\lambda_n - \lambda_1)\mu_1 = (\lambda_n - \lambda_2)\mu_2 = \ldots = (\lambda_n - \lambda_{n-1})\mu_{n-1} = 0$$

are all zero. Since the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all distinct, this must imply that

$$\mu_1 = \mu_2 = \ldots = \mu_{n-1} = 0$$

Hence

$$\mu_n v_n = \mu_1 v_1 + \mu_2 v_2 + \ldots \mu_n v_n = 0$$

So, since $v_n \neq 0$, $\mu_n = 0$. So $v_1, \ldots, v_n$ are linearly independent. □

**(7.9) Corollary.**
*Suppose that $V$ is an $n$-dimensional vector space over $\mathbb{K}$ and that $f : V \to V$ is a linear mapping. If $f$ has $n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $V$ has a basis of eigenvectors.*

*Proof.* Let $v_1, v_2, \ldots, v_n$ be eigenvectors associated with $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then $v_1, \ldots v_n$ are linearly independent and span a subspace of $V$ of dimension $n$. Since $V$ has dimension $n$, this is the whole of $V$. So $\{v_1, v_2, \ldots, v_n\}$ is a basis for $V$. □

**(7.10) Corollary.**
*Suppose that $A$ is an $n \times n$ matrix over a field $\mathbb{K}$, with $n$ distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in k$. Then $A$ is diagonalisable*

*Proof.* This is a direct consequence of the theorem and the proposition directly preceding it. □

**(7.11) Exercise.**
Find a $2 \times 2$ matrix $B$ such that the matrix $B^{-1}AB$, where

$$A = \begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix},$$

is diagonal.

**(7.12) Solution.**

$$\begin{aligned}
\det(A-\lambda I_2) &= \begin{vmatrix} 1-\lambda & 3 \\ 3 & -1-\lambda \end{vmatrix} = (1-\lambda)(-1-\lambda)-9 \\
&= -1+\lambda^2-9 = \lambda^2-10,\ \lambda = \pm\sqrt{10} \\
(A-\sqrt{10}I_2)\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\leftrightarrow (1-\sqrt{10})x+3y &= 0,\ 3x-(1+\sqrt{10})y = 0, \\
\leftrightarrow \begin{pmatrix} x \\ y \end{pmatrix} &= k\begin{pmatrix} 3 \\ \sqrt{10}-1 \end{pmatrix} \\
(A+\sqrt{10}I_2)\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\leftrightarrow (1+\sqrt{10})x+3y &= 0,\ 3x+(\sqrt{10}-1)y = 0, \\
\leftrightarrow \begin{pmatrix} x \\ y \end{pmatrix} &= k'\begin{pmatrix} \sqrt{10}-1 \\ -3 \end{pmatrix} \\
B &= \begin{pmatrix} 3 & \sqrt{10}-1 \\ \sqrt{10}-1 & -3 \end{pmatrix}
\end{aligned}$$

A matrix with less than $n$ distinct eigenvalues might or might not be diagonalisable.

**(7.13) Examples.**
The following matrices have repeated eigenvalues and <u>are</u> diagonalisable:-

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

The first is diagonal, so is certainly diagonalisable. We see that they are diagonalisable by producing matrices which diagonalise.
A second example is the matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

This matrix has characteristic polynomial $X^2+1$. This means that it has no eigenvectors over $\mathbb{R}$ so over $\mathbb{R}$ it is not diagonalizable. However over $\mathbb{C}$ it is diagonalisable: its eigenvectors are $(1,i)$ with eigenvalue $i$ and $(1,-i)$ eigenvalue $-i$.
Even over $\mathbb{C}$ not all matrices are not diagonalisable. The following matrices have all eigenvalues (over $\mathbb{R}$) and <u>are not</u> diagonalisable over $\mathbb{R}$ and over

$\mathbb{C}$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

For if they were diagonalisable, in each case the eigenvectors with eigenvalue 1 would span the space. So the matrix would be a matrix of the identity mapping, and hence would be the identity matrix.
In fact the same kind of argument shows that any upper triangular or lower triangular matrix with all diagonal entries the same is diagonalisable if and only if it is already diagonal.

# 2 The Jordan normal form

Now we are going to investigate in what form we can bring a matrix if it is not diagonalisable over $\mathbb{C}$ (any other algebraicly closed field).

**(7.14) Definition.**
Let $V$ be a vector space and $f : V \to V$ a linear map. We call a sequence of nonzero vectors $v_1, \ldots, v_k$ a Jordan sequence with eigenvalue $\lambda$ if

$$\forall i \in [1, k-1] v_{i+1} = f(v_i) - \lambda v_i \text{ and } f(v_k) - \lambda v_k = 0$$

Note that Jordan sequences always exist because every eigenvector forms a Jordan sequence of length 1.

**(7.15) Lemma.**
*If $v_1, \ldots, v_k$ is a Jordan sequence with eigenvalue $\lambda$ then $v_1, \ldots, v_k$ are linearly independent and $f$ will map $\langle v_1, \ldots, v_k \rangle$ to itself. We can express $f|_{\langle v_1, \ldots, v_k \rangle}$ as the matrix*

$$\mathcal{J}(\lambda, k) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

*This matrix is called a Jordan block*

*Proof.* Suppose $\sum a_i v_i = 0$ and $j$ is the first coefficient such that $a_j \neq 0$. Then $(f - \lambda \mathbb{I})^{k-j}(\sum a_i v_i) = a_j v_k = 0 \implies v_k - 0$. The rest of the lemma follows from the definition of a Jordan sequence. □

**(7.16) Theorem.**
*Every linear map $f : V \to V$ with $V$ a finite dimensional vector space over $\mathbb{C}$ has a basis consisting of Jordan sequences (a.k.a a Jordan basis).*

*Proof.* Suppose that the statement is false, then there exists a $V, f$ that has no basis consisting of Jordan sequences. We take the $V, f$ with the smallest possible dimension. This dimension is a least 2 (because every basis for a one-dimensional space consists of an eigenvector).
Because we are working over the complex numbers $f$ has at least one eigenvector. Without loss of generality, we can assume that $f$ has an eigenvector with eigenvalue 0, because a Jordan basis for $V, f$ is also a Jordan basis for $V, f - \lambda$ Therefore $\operatorname{Ker} f \neq 0$ and hence $\operatorname{Im} f < \dim V$. Clearly $f(\operatorname{Im} f) \subset \operatorname{Im} f$ so we can consider the map

$$f|_{\operatorname{Im} f} : \operatorname{Im} f \to \operatorname{Im} f : u \longmapsto f(u)$$

The pair $\operatorname{Im} f, f|_{\operatorname{Im} f}$ has a strictly smaller dimension so it has a Jordan basis. Now look at the set of subspaces $U \subset V$ with $f(U) \subset U$ that that have a Jordan basis which contains a basis of $\operatorname{Im} f$.
This set is not empty as $\operatorname{Im} f$ itself is in it. We take an element of this set with maximal dimension and call it $W$. Its Jordan basis will be denoted by $B$.
We will prove that $W$ must be equal to $V$.
First we show that $f(W) = \operatorname{Im} f = f(V)$. Clearly as $W \subset V$, $f(W) \subset f(V)$ so we have to prove that $f(V) \subset f(W)$. It suffices to show that $B \cap f(V) \subset f(W)$ because $B$ contains a basis for $f(V)$.
Let $w \in B \cap f(V)$ then it is contained in a Jordan sequence $v_1, \ldots, v_k \in W$ with eigenvalue $\lambda$.

$\lambda = 0$ If $w$ is not the first element of the sequence, it is equal to $f(v_i) \in f(W)$ for some $v_i$. If $w = v_1$ then we can find a $v_0$ such that $f(v_0) = w$ because $w \in f(V)$. If $w' \notin W$ then $B \cup v_0$ is a Jordan basis for $\langle W, v_0 \rangle$ ($v_0, v_1, \ldots, v_k$ is a Jordan sequence) which contradicts the fact that $W$ was maximal.

$\lambda \neq 0$ If $v \in W$ then $v \in f(W) \iff -\lambda v + f(v) = (f(v) - \lambda \operatorname{id})v \in f(W)$. So for a Jordan sequence we get $v_i \in W \iff v_{i+1} \in W$ and as $(f(v) - \lambda \operatorname{id})v_k = 0 \in f(W)$ we get that the every Jordan sequence with $\lambda \neq 0$ sits in $f(W)$.

Because $f(W) = f(V)$, we can find for any $v \in V$ a $w \in W$ with $f(v) = f(w)$. As $f(v - w) = 0$ the vector $v - w$ forms a Jordan sequence of length 1. If $v - w \notin W$ then $B \cup \{v - w\}$ forms a Jordan basis for $\langle W, v - w \rangle$ containing a basis for $f(V)$. This contradicts the fact that $W$ is maximal. So $v - w \in W$ and hence $v \in W$, so $V = W$. □

**(7.17) Theorem.**
*Every matrix $A$ can be conjugated to a direct sum of Jordan blocks.*

$$A \cong_C \mathcal{J}(\lambda_1, k_1) \oplus \cdots \oplus \mathcal{J}(\lambda_p, k_p)$$

*Two direct sums of Jordan blocks are conjugate if they have the same number of blocks of each type. I.e.*

$$\mathcal{J}(\lambda_1, k_1) \oplus \cdots \oplus \mathcal{J}(\lambda_p, k_p) \cong_C \mathcal{J}(\mu_1, l_1) \oplus \cdots \oplus \mathcal{J}(\mu_q, l_q)$$

*if and only if $p = q$ and $\exists \pi \in \mathrm{Perm}(\{1, \ldots, n\}) : \lambda_i = \mu_{\pi(i)} \& k_i = l_{\pi(i)}$.*

*Proof.* The first statement is a reformulation of the previous theorem.
The second requires a bit more thought. Clearly the condition is sufficient because $A \oplus B \cong_C B \oplus A$.
It is also necessary because we can express the $k_i$ and $\lambda_i$ in terms of the linear map representing the matrix.
Indeed $\mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)$ equals the number of jordan blocks with eigenvalue $\lambda$, while $\mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)^2 - \mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)$ equals the number of jordan blocks with eigenvalue $\lambda$ and size $k \geq 2$. In general $\mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)^i - \mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)^{i-1}$ equals the number of jordan blocks with eigenvalue $\lambda$ and size $k \geq i$.
As $\mathrm{Dim\ Ker}(BAB^{-1} - \lambda \mathbb{I}_n)^i = \mathrm{Dim\ Ker}\, B(A - \lambda \mathbb{I}_n)^i B^{-1} = \mathrm{Dim\ Ker}(A - \lambda \mathbb{I}_n)^i$ The number of Jordan blocks of each type in the two direct sums must be the same. □

**(7.18) Definition.**
A square matrix is called indecomposable if it cannot be conjugated to a direct sum of smaller matrices.

**(7.19) Theorem.**
*Any indecomposable matrix can be conjugated to a Jordan block.*

# 3 The characteristic polynomial

If $A \in \mathrm{Mat}_{n \times n}(\mathbb{K})$ is a square matrix and $p(X) = p_0 + p_1 X + \ldots p_k X^k \in \mathbb{K}[X]$ is a polynomial, we can calculate the matrix

$$p(A) := p_0 \mathbb{I}_n + p_1 A + \cdots + p_k A^k.$$

**(7.20) Lemma.**

- $p(GAG^{-1}) = Gp(A)G^{-1}$.
- $p(A \oplus B) = p(A) \oplus p(B)$

*Proof.* First note that $(GAG^{-1})^i = (GAG^{-1})(GAG^{-1})\dots(GAG^{-1}) = GAA\dots AG^{-1} = GA^iG^{-1}$ and $\mathbb{I}_n = GG^{-1}$. As a consequence

$$p(GAG^{-1}) = p_0GG^{-1} + p_1GAG^{-1} + \dots + p_kGA^kG^{-1} = Gp(A)G^{-1}.$$

The second fact follows from the following easy identities: $\forall A_1, A_2 \in \mathrm{Mat}_{n\times n}$ : $\forall B_1, B_2 \in \mathrm{Mat}_{m\times m}$ :

$$\begin{aligned}(A_1 \oplus B_1) + (A_2 \oplus B_2) &= (A_1 + A_2) \oplus (B_1 + B_2)\\ (A_1 \oplus B_1)(A_2 \oplus B_2) &= (A_1A_2) \oplus (B_1B_2)\\ \mathbb{I}_n \oplus \mathbb{I}_m &= \mathbb{I}_{n+m}.\end{aligned}$$

□

**(7.21) Definition.**
We call the polynomial $\chi_A(X) := \det(A - X\mathbb{I}_n)$ the characteristic polynomial of $A$. The eigenvalues of $A$ are the roots of $\chi_A(X)$.

**(7.22) Lemma.**
- *Conjugate matrices have the same characteristic polynomial.*

$$\chi_{GAG^{-1}}(X) = \chi_A(X)$$

- *The characteristic polynomial of the direct sum of two matrices is the product of their characteristic polynomials.*

$$\chi_{A\oplus B}(X) = \chi_A(X)\chi_B(X)$$

*Proof.* $\det(GAG^{-1} - X\mathbb{I}_n) = \det(GAG^{-1} - XGG^{-1}) = \det(G(A - X\mathbb{I}_n)G^{-1}) = \det(G)\det(A - X\mathbb{I}_n)\det(G^{-1}) = \det(A - X\mathbb{I}_n)$.
The second identity follows from the fact that $\det(A\oplus B) = \det(A)\times\det(B)$.
□

We are now ready to prove a very important theorem:

**(7.23) Theorem. Cayley-Hamilton Identity**
*If you fill in $A$ in its characteristic polynomial you get the zero matrix.*

$$\chi_A(A) = 0$$

*Proof.* Because of the lemmas above we only need to show this for Jordan blocks. One can easily check that $\chi_{\mathcal{J}(\lambda,k)}(X) = (X-\lambda)^k$. Now $\mathcal{J}(\lambda,k)-\lambda\mathbb{I}_k$ a maps $e_k \longmapsto e_{k-1} \longmapsto \ldots \longmapsto e_1 \longmapsto 0$ so $(\mathcal{J}(\lambda,k) - \lambda\mathbb{I}_k)^k$ is the zero map. □